Р2008
1001ұ

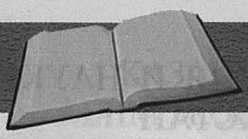
Аскар ЖҰБАНОВ

ҚОЛДАНБАЛЫ ТІЛ БІЛІМІНІҢ МӘСЕЛЕЛЕРІ

ВОПРОСЫ ПРИКЛАДНОГО ЯЗЫКОЗНАНИЯ



P2008/1001K



Тіл білімінің озық үлгілері

АРЫС

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН
ИНСТИТУТ ЯЗЫКОЗНАНИЯ им. А.БАЙТУРСЫНОВА

МИНИСТЕРСТВО КУЛЬТУРЫ И ИНФОРМАЦИИ
РЕСПУБЛИКИ КАЗАХСТАН
КОМИТЕТ ПО ЯЗЫКАМ

Аскар ЖУБАНОВ

ВОПРОСЫ ПРИКЛАДНОГО ЯЗЫКОЗНАНИЯ

Издательство "Арыс"
Алматы
2008

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
А.БАЙТҰРСЫНҰЛЫ АТЫНДАҒЫ ТІЛ БІЛІМІ ИНСТИТУТЫ

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
МӘДЕНИЕТ ЖӘНЕ АҚПАРАТ МИНИСТРЛІГІ
ТІЛ КОМИТЕТІ

Асқар ЖҰБАНОВ

ҚОЛДАНБАЛЫ ТІЛ БІЛІМІНІҢ МӘСЕЛЕЛЕРІ

"Арыс" баспасы
Алматы
2008

ББК 81.2-5
Ж 80

Кітап "Арыс" қорының өз қаражатымен "Қазақ ғылымының озық үлгілері" атты мегажобасының аясында жарық көріп отыр.

Книга выпущена фондом "Арыс" на собственные средства в рамках мегапроекта "Лучшие образцы казахской науки".

Ж 80 **ЖҰБАНОВ А.**
Қолданбалы тіл білімінің мәселелері. Вопросы прикладного языкознания. – Алматы: "Арыс" баспасы, 2008. – 640 бет. + 8 бет суретті жапсырма.

ISBN 9965-17-460-1

Қолданбалы қазақ тіл білімінде зор еңбек етіп келе жатқан тілші-математик Асқар Жұбановтың зерттеулер жинағында статолингвистика, математикалық және компьютерлік лингвистика салаларының әдіс-тәсілдері мен қыр-сырлары, бүгіні мен болашағы жан-жақты сараланып, филолог мамандар үшін түсінікті түрде баяндалады.

Кітап жоғары оқу орындарының оқытушыларына, аспиранттар мен ізденушілерге, магистранттар мен студенттерге және жалпы көпшілік қауымға арналған.

Лингвист-математик Аскар Жубанов вносит большой вклад в развитие прикладной лингвистики казахского языкознания. В данной книге собраны основные труды автора, где разрабатываются методология статолингвистической, математической и компьютерной лингвистики, исследуются различные аспекты данной отрасли языкознания, его современное состояние и перспективы развития.

Книга излагается в доступной для специалистов-филологов форме и адресована преподавателям высших учебных заведений, магистрантам и студентам, а также для всех интересующихся читателей.

Ж $\frac{44602020400}{00(05)-07}$ ББК 81.2-5

ISBN 9965-17-460-1

© Жұбанов А., 2008
© "Арыс", 2008

Осы еңбегімді қазақ тіл білімінің
классигі, ойшыл-филолог ғалым,
1937 жылғы нәубеттің құрбаны –
Әкем Құдайберген Қуанұлы ЖҰБАНОВТЫҢ
ұлылығына бас ие отырып,
әруағына бағыштадым...

Преклоняясь перед гениальностью
Отца Кудайбергена Куановича ЖУБАНОВА –
филолога-мыслителя, классика казахского
языкознания, ставшего жертвой
репрессий 1937 года, посвящаю
свой скромный труд...

Автор

ҚАЗАҚ ТІЛТАНЫМЫНЫҢ ЖАҢА БІР САЛАСЫНЫҢ КӨШБАСТАРЫ

Асқар Жұбанов 1937 жылдың 13 қарашасында Алматы қаласында қазақ тіл білімінің теориялық негізін салушы атақты ғалым, тұңғыш кәсіби лингвист, түркітанушы, ағартушы, профессор Құдайберген Жұбановтың отбасында дүниеге келді. Бұл бүгінде баршамызға белгілі саяси қуғын-сүргіннің нағыз өршіп тұрған кезі болатын. Бір аптадан соң репрессия өрті ғалымды да жалмады, "жапон шпионы" деген жаламен түрмеге қамалып, артынша миллиондаған сталиндік режимнің құрбандарының жолын құшты...

Шыр етіп жарық дүниенің есігін ашқаннан тағдырдың тепкісін көрген кенже бала Асқар өзінің екі аға, үш апасы, "халық жауының әйелі" атанған анасымен бірге Қазақстанның бір қиырынан екінші шетіне қоныс аударып, көшіп жүруге мәжбүр болды. Ата жұрт – Ақтөбе өңірі, Алматы, солтүстіктегі Петропавл, қайтадан Алматы деп, қашан ардақты әке есімі "ақталғанша" бір жерде тұрақтатпады. Алдымен Құдекеңнің – Құдайберген Жұбановтың әкелік мейірімі мен тәрбиесін, Ахаңның – академик Ахмет Жұбановтың қамқорлығы мен көмегін, жақсылығын көріп өскен, қала берді, қанға біткен қайсарлықтың арқасында әкеден шиеттей жетім қалған алты сарыауыз балапанның жерде қалғаны жоқ: отыз жылдай Алматыдағы сурет училищесіне басшылық еткен Ақырап, профессор Есет,

---

ОСНОВАТЕЛЬ НОВОГО НАПРАВЛЕНИЯ КАЗАХСКОГО ЯЗЫКОЗНАНИЯ

Аскар Жубанов родился в семье известного ученого, основателя теории казахского языкознания, первого профессионального лингвиста, тюрколога, просветителя, профессора Кудайбергена Куановича Жубанова 13 ноября 1937 г. в городе Алматы. Это было время, как известно нам из истории, свирепствования политической репрессии. Через неделю волна репрессии подхватила и ученого, по обвинению "японский шпион", его посадили в тюрьму, после его настигла та же участь, что и других миллионных жертв сталинского режима.

Не успев появиться на свет, младший сын Аскар вместе с двумя старшими братьями, тремя сестрами и матерью, за которой прикрепили обвинение "жена врага народа", были вынуждены скитаться с одного угла Казахстана на другой. До "оправдания" доброго имени отца они вынуждены были не задерживаться ни на одном месте: родной край – Актюбинск, Алматы, северный Петропавловск, затем снова Алматы. Успевшие повидать отцовское воспитание и нежность Құдекең – Кудайбергена Жубанова, испытать заботу и опеку Ахаң – Ахмета Жубанова, но рано потерявшие отца дети, несмотря на все тяготы жизни и трудности, выпавшую на их долю, стали достойными людьми: имя художника Акырап, профессора Есет, автора "Букваря – Алиппе", педагога Муслимы,

"Әліппенің" авторы, ұстаз Мұслима, экономист Қызғалдақ, Қырмызылар есімі еліміздің тарихымен бірге өріліп жатыр. Солардың бірі – бүгінде әке ісін жалғастырып, қазақ тіл білімінің тың саласының түренін тартып, тізгінін ұстап отырған профессор Асқар Жұбанов.

Мектеп қабырғасында нақты пәндерді ұнатқан жас Асқар оқу бітірген соң Абай атындағы Қазақтың педагогикалық институтының физика-математика факультетіне оқуға түсіп, 1960 жылы оны математика мамандығы бойынша үздік бітіріп шықты. Одан соңғы аралықтарда әскерлік борышын өтеп келіп, табан аудармай Қазақ мемлекеттік ауылшаруашылық институтының жоғары математика кафедрасында қызмет атқарады. Күндердің күнінде, отызға толған шағында көп ойланып, көп толғанып жаңа жолға, ғылымның соны соқпағына бет бұрады.

Бұл жылдар советтік тарихымызда "жылымық жылдар" атанып, ел еркін тыныстап, әдебиет пен өнер, ғылым мен техника жаңа сапада, тың серпінмен дами бастаған кезі еді. Дәстүрлі, классикалық ғылымдардың өзі жаңа өріс ізденп, пәнаралық, салааралық бұлақ көздері ашыла бастады. Сондай жаңа саланың бірі, өкінішке орай, тап бүгіннің өзінде қазақ топырағында өркенін кең жая алмай, көпшілік тосырқана қарап келе жатқан математикалық лингвистика еді. Бірде статолингвистика, бірде компьютерлік лингвистика, кейде машиналық аударма сияқты аттарға ие болып, "қолданбалы лингвистика" атты ортақ шапан киген пәнаралық ғылымға бәріміз шекеміздің қарап жүргенде бұл жасанды интеллектіңіз жекебас компьютері түрінде қағаз-қаламды айырбастап, бәріміздің төріміздің бір-ақ шықканын да байқамай қалдық.

экономиста Кызгалдак и Кырмызы, связано с историей нашего народа. Один из них – Аскар Жубанов продолжает дело отца, он занимается новым направлением в казахском языкознании.

Полюбив еще со школьной скамьи точные предметы, Аскар после окончания школы поступил на физико-математический факультет Казахского педагогического института имени Абая и в 1960 году успешно завершил его по специальности математика. Затем, отслужив в рядах армии, работал на кафедре высшей математики Казахского государственного сельскохозяйственного института. К тридцати годам он вышел на новый путь науки.

Этот советский период нашей истории считался "периодом оттепеля", когда народ мог свободно дышать, литература и искусство, наука и техника стали развиваться в разных направлениях. Традиционные классические науки стали искать новые междисциплинарные, межотраслевые подходы. Одним из этих направлений, к сожалению, не нашедших в данный момент широкого применения, является математическая лингвистика. В начале междисциплинарная наука получившая название статолингвистика, машинный перевод, затем компьютерная лингвистика, объединенная под общим названием "прикладная лингвистика" обрела почетное место, связано было это заменой пера и бумаги на искусственный интеллект – компьютер.

В эти годы в Москве и Ленинграде, а также в Минске стали создаваться исследовательские центры, связывающие лингвистику с математикой.

Осы жылдар Мәскеу мен Ленинградта, сондай-ақ Минскіде лингвистика мен математиканы ұштастыра зерттейтін орталықтар құрыла бастады. Оның бастауында профессор Р.Г.Пиотровский тұрған еді. Қашанда ғылымда жаңа нәрсе тез таралады. Қазақстанда шымкенттік математик (кейіннен тілші ғалымға айналды) Қалдыбай Бектаев осы іске бел шеше кірісті. Болашақты болжай білген Қазақ ССР Ғылым академиясының басшылары, оның ішінде Тіл білімі институтының директоры Ісмет ага Кеңесбаев уақтылы осы бастамаға үн қоса білді, ең алғашқы одақтас республикалардың бірі болып Академиялық институт жанынан арнайы ғылыми топ құрылды. Міне, осы жаңа істің басы-қасында алғашқылардың бірі болып "пионер" Асқар Жұбанов та жүрді.

Әрине, математик Асқардың тілшілік (математикалық-тілшілік, болмаса тілшілік-математикалық) жолды таңдауы оңай болмады. Алдымен әке жолын қуып туабітті тілші һәм әдебиетші бола білген, сөзі ұзын, қаламы жүйрік Есеттей ағасының үгіт-насихаты бір себепші болса, екіншіден, көзі көрмесе де құлағы қанық ғалым әкесін арбаған тіл ғылымының тұңғиығына өзі де ең болмағанда бір рет сүңгіп шығуға аңсары ауғаны кәдік. Сөйтіп, 1967–1970 жылдары қазақ тілін кешенді зерттейтін әлемдегі жалғыз қара шаңырақ – Қазақ ССР Ғылым академиясының Тіл білімі институтының күндізгі бөлімінің аспиранты болады және осы үш жылын одақтық деңгейдегі математикалық лингвистиканың үлкен орталығы – Минск қаласындағы Шет тілдер институтында өткізді. Осы жерде өзі секілді аспирант, бірақ кейін ғылым жолында ұстазындай бола білген А.В.Зубовпен табысып, күні бүгінге дейін араласып, ақылдасып тұрады. Ол жылдарда

Основателем был профессор Р.Г. Пиотровский. В науке новое всегда быстро находит распространение. В Казахстане известный математик-вероятностник (после стал языковедом) Калдыбай Бектаев вплотную стал заниматься этим направлением. Сумевшие предвидеть перспективу данного направления руководство Академии наук КазССР, среди них директор Института языкознания Смет ага Кенесбаев вовремя поддержало это начинание, среди союзных республик первыми организовали специальную научную группу при Академическом институте. Среди всех принимал в этом участие и "пионер" Аскар Жубанов.

Конечно, для математика Аскара выбрать путь языковеда было не легко. Но несмотря на это, во-первых, было влияние брата Есета, который сумел стать хорошим языковедом и литератором, как и отец, во-вторых, он сам также мечтал окунуться в науку, которой занимался его отец. Таким образом, в 1967–1970-е годы он становится аспирантом дневного отделения единственного в мире очага, занимавшегося исследованием казахского языка – Института языкознания Академии наук КазССР и эти три года он провел в Институте иностранных языков в городе Минске – большом математико-лингвистическом центре на уровне союза. Здесь он знакомится с таким же аспирантом, после ставшим учителем, А.В. Зубовым и до сих пор поддерживает с ним связь. В те годы основным средством статолингвистики был ЭВМ (электронно-вычислительная машина), который, естественно, "не говорил" на казахском. Аскар Жубанов, как

статолингвистердің негізгі құралы – ЭЕМ (электронды есептеуіш машина), оның өзі қазақша "сөйлеп" көрген емес. Оған арнайы бағдарлама түзіп, түрлі сөз табына жататын сөздерді, сөз қолданыстарды машина көмегімен анықтап, неше түрлі жиілік, кері алфавиттік сөздіктер түзу сияқты тың істердің басында тұрған, тура мағынасындағы "пионер" ғалым Асқар Жұбанов болатын. Осылай тыңға түрен сала жүріп ол 1973 жылы кандидаттық жұмысын табысты қорғады және осы жолдағы ұзақ жылғы ізденістерін "Квантативная структура казахского текста" деген атпен ("Ғылым", 1987) жарыққа шығарды. Осылайша қазақ ғалымы қазақ тілі мәтіндерін ЭЕМ-нің жасанды миында "қайнатып" алып, алғаш рет ана тіліміздегі тілдік бірліктердің (сөзформа, негіз сөз, әріп, әріп тіркесі т.б.) автоматтанған жиілік сөздіктерін құрастырудың алгоритмі мен бағдарламасын жасап шығып, оны іске қосты. Сонымен бірге қазақ мәтініне сүйене отырып жалпы түркі тілдері мәтіндерінің ықтималдық-статистикалық моделін жасап шығудың алғашқы нәтижелеріне қол жеткізілді. Математикалық статистика мамандары ғана түсіне алатын Нормаль, Пуассон, Шарлье үлестірілімдік заңдарына сүйене отырып, қазақ мәтініндегі зат есім мен етістіктердің алғашқы заңға бағынатынын, көпшілік жағдайда үстеу сөздердің Пуассон заңы мен Шарлье заңының В түріне байлаулы екеніне алғаш рет көз жеткізілді. Осылайша қазақ мәтініндегі мыңдаған сөзқолданыстардың бір қазанда кездейсоқ бас қосып, қайнаған материалдар емес, сөз таптарына ажыратылу, үлес салмағы тұрғысынан белгілі бір ықтималдық-статистикалық заңдылыққа бағынатыны дәлелденді.

истинный "пионер" этого дела, стоял у истоков создания специальной программы, которая дает возможность с помощью ЭВМ создавать различные частотные словари, как обратно-алфавитные частотные словари, которые дают возможность определить слова, относящиеся к разным частям речи. Плодотворно трудясь, он в 1973 году успешно защитил кандидатскую диссертацию и выпустил на свет свой труд "Квантативная структура казахского текста" ("Наука", 1987). Таким образом, казахский ученый ввел в память ЭВМ тексты, относящиеся к различным стилям казахского языка и впервые составил на ЭВМ алфавитно-частотные словари различных языковых единиц (словоформа, основа, буква, сочетание букв и др.). Вместе с тем, основываясь на полученные данные по текстам казахского языка, смог добиться определенных результатов по созданию вероятностно-статистических моделей тюркских текстов. Опираясь на Нормальный закон распределения, на законы распределения Пуассона и Шарлье, понятные только математикам, впервые определил, что в казахском тексте имя существительное и глагол подчиняются Нормальному закону, а наречие в основном подчиняется законам Пуассона и Шарлье типа-В. Тем самым было доказано, что использование слов в казахском тексте это не просто набор слов, а совокупность текстовых единиц, распределенных в тексте, подчиняясь определенным вероятностно-статистическим законам.

В своей кандидатской диссертации ученый определяет статистическую структуру эпопеи "Путь Абая" М. Ауэзова. После, трудясь плодотворно в этом

Ғалым осы кандидаттық жұмысының өзінде М.Әуезовтің "Абай жолы" эпопеясының статистикалық құрылымын анықтауға ұмтылды. Кейін осы салада өзімен бірге өнімді еңбек еткен ғалымдармен бірлесе отырып "М.Әуезовтің "Абай жолы" романының жиілік сөздігін" түзіп шықты ("Ғылым", 1979, 336 бет). Осы еңбектің нәтижесінде классик жазушымыздың эпопеяда 465691 сөзқолданысты пайдаланғанын, ондағы сөз саны – 16983, ал сөзформасы – 61424 екені анықталды. Сонымен бірге белгілі түркітанушы ғалым Ә.Құрышжановпен бірлесе отырып жарыққа шығарған "Куманша-қазақша жиілік сөздік" те ("Ғылым", 1978, 216 бет) түсіне білген жанға сандық көрсеткіш арқылы тілдік сапаны жарқырата ашып көрсететін күрделі еңбек-тұғын. Өйткені әлі күнге ана тілімізге толықтай аударылмай әрі асыл нұсқасы жарық көре алмай келе жатқан "Кодекс Куманикс" (XIII г.) – қазақ тіліне тікелей қатысты көне қыпшақ жәдігері.

Ғалым 1997 жылы тағы бір жыл бойы Белоруссия Ғылым академиясының Тіл білімі институтында докторантура есебіндегі ғылыми іссапарда болады. Профессор А.В.Зубовтың кеңесшілігімен ұзақ жылдарғы зерттеуінің қорытындысы есепті "Основные принципы формализации содержания казахского текста" деген тақырыпта монографиясын жарыққа шығарып (Алматы, 2002, 250 бет), докторлық диссертациясын сәтті қорғап шықты. Тіл білімінде соңғы жылдары ғана алдыңғы қатарға шыққан жаңа зерттеу бағыты – мәтін лингвистикасы тұрғысынан жазылған, мәтіннің мазмұндық құрылымының түзілуін абзацтық деңгейде қарастырған күрделі еңбектің негізгі мақсаты – қазақ мәтінінің семантикалық көрінісін бейнелеу және оның туындау мүмкіндігінің формалды принциптерін анықтап, көрсе-

---

направлении, вместе с другими учеными, создал "Частотный словарь романа М. Ауэзова "Путь Абая" ("Наука", 1979, 336 с.). В результате этого труда удалось определить, что писатель-классик использовал в эпопее 465691 словоупотреблений, из них 16983 – слова, а словоформы – 61424. Кроме того, созданный совместно с известным тюркологом, ученым А. Курышжановым "Частотный куманско-казахский словарь" ("Наука", 1978, 216 с.) является сложным трудом, который может раскрыть качественное использование языковых единиц с помощью статистических данных. Этот труд является важным и ценным, так как до сих пор полностью не переведенный на казахский язык "Кодекс Куманикс" (XIII в.) – кыпчакский памятник, имеет прямое отношение к казахскому языку.

Ученый в 1997 году был в командировке по докторантуре в Институте языкознания Академии наук Белоруссии. В результате полученных консультаций от профессора А.В.Зубова выпустил в свет свой труд "Основные принципы формализации содержания казахского текста" (Алматы, 2002, 250 с.), как итог многолетней работы, успешно защитил докторскую диссертацию. Основная цель ценного труда, который рассматривает содержательную структуру текста на уровне абзаца, написанный по новому направлению языкознания – лингвистике текста – описание семантической картины казахского текста и определить, выделить формальные принципы их порождения. Таким образом, мы можем смело сказать, что это были попытки применить современные информационные технологии для расширения рамки использования государственного языка. Предлагаемая ученым

ту. Бір сөзбен айтқанда, бұл мемлекеттік мәртебеге ие болып отырған ана тіліміздің қазіргі заманғы ақпараттық технологиялардың жаңалықтарына жедел бой үйретіп, соның нәтижесінде қолданыс аясының кеңеюіне ықпалдастық жасау. Атап айтқанда, зерттеуші ұсынып отырған ана тіліміздің формалданған моделі машиналық аударма тәжірибесінде, тіл оқытудың компьютерлік бағдарламасын жасауда, қазақ лексикографиясын жаңа деңгейге көтеруге, яғни оны автоматтандыруға және т.б. тек теориялық тұрғыдан ғана емес, осындай қыруар қолданбалы, практикалық мәселелерді шешуде де пайдасы шаш етектен болмақ.

Міне, осылайша осыдан 40 жыл бұрын тілшілер ауылына қорқасоқтап, қымсына аяқ басқан талапты жас бүгінде қазақ тіл білімінің жаңа да болашағы зор "қолданбалы лингвистика" саласының көшбастаушы "ақсақалы" дәрежесіне абыроймен жетіп отыр. "Шөп те шыққан жеріне шығады", 40 жылдан бері әке жолын жалғап, табан аудармай А.Байтұрсынұлы атындағы Тіл білімі институтында қызмет атқарып келеді. Аспирант, кіші ғылыми қызметкер, 1976 жылдан аға ғылыми қызметкер, 1986 жылдан жетекші ғылыми қызметкер, ұзақ жыл "Қолданбалы лингвистика" бөлімін басқарды (1990–97). Соңғы онжылдықтағы түрлі реформа, өзгерістерге байланысты Институтта өз саласы бойынша бас ғылыми маман ретінде қызмет атқарып келеді.

А. Жұбановтың тікелей атсалысуымен Институтта соңғы жылдары атқарылып келе жатқан үлкен жоба – "Қазақ тілінің компьютерлік қоры". Мұнда ана тіліміздің бар байлығы, ақпараттанған білім қоры жинақталған. Соңғы кезде "Тіл–Қазына" атауына ие болған картотекалық қордың компьютерлік базасын асыл

формальная модель родного языка может быть использована в переводческой деятельности, в создании компьютерной программы по обучению языку, для автоматизации задач, относящихся к лексикографической отрасли казахского языкознания, а также в решении не только теоретических, но и ряда практических проблем.

Вступивший 40 лет назад на путь языковеда молодой ученый на сегодняшний день справедливо занял почетное место "аксакала" среди основоположников "прикладной лингвистики". В течение 40 лет по стопам отца он преданно служит в Институте языкознания имени А.Байтурсынова: аспирант, младший научный сотрудник, с 1976 года старший научный сотрудник, с 1986 года ведущий научный сотрудник, долгие годы руководил отделом "Прикладная лингвистика" (1990–97). В последние десятилетия в связи с различными реформами в Институте работает как главный научный сотрудник.

По инициативе А. Жубанова в Институте в последние годы работают над большим проектом – "Компьютерная база казахского языка". Здесь собрано все богатство, информационный фонд языка. Получившее название "Тіл – Қазына" – картотечный фонд компьютерной базы можно назвать "Банком данных о казахском языке", где хранятся все богатство родного языка, культурное наследие. В будущем новая волна молодых ученых, которые в совершенстве будут владеть компьютером, смогут использовать это богатство и внесут большой вклад в развитие казахской лексикографии.

сөз, мәдени мұрамыз сақталатын "Тілбанкі" десек болғандай... Таяу болашақта компьютердің құлағында ойнайтын жаңа толқын ғалымдар ол байлықты сарқа пайдаланары, әсіресе кенжелеп қалған лексикографиямыздың өркендеуіне қосар үлесі қомақты болары сөзсіз.

Профессор А. Жубановты соңғы жылдары Алматыдағы жетекші білім ошақтары – әл-Фараби атындағы Қазақ ұлттық университеті мен Абай атындағы Алматы ұлттық педагогикалық университеттері аттай қалап шақырып, қолданбалы лингвистиканың түрлі саласынан арнайы пәндер ашып, студенттеріне дәріс тыңдатып жүр. "Ұстаздық еткен жалықпас" дегендей, бұл орайда да ғалым ісі өнікті. Сол жерден өзіне бірнеше ізбасар, шәкірт тапты, алды кандидатттық диссертацияларын қорғап, кейбірі ғылым жолына, тың салаға атбасын бұрып жатыр. Бұған дейін басты еңбектерін орыс тілінде жазып келе жатқан А. Жубанов осы университет қабырғасында оқыған дәрістері негізінде қолданбалы лингвистиканы ана тілімізде сөйлетіп, оқулықтардың тұтас сериясын жасап шығаруды мақсат тұтқандай. Оған мына қолымызға ұстап отырған оқу құралдары куә: "Компьютерлік лингвистикаға кіріспе" (Алматы, 2007), "Қолданбалы лингвистика: қазақ тілінің статистикасы" (2-бас., Алматы, 2007), "Қолданбалы лингвистика: формалды модельдер" (2-бас., Алматы, 2007).

Адам өзі өрекпіп көрсетпесе де, даурығып дәлелдемесе де, ең мықты ғалым оның бойында жатқан небір жақсы қасиеттері барлардан шығатын сияқты. Жубановтар әулетінің қай өкіліне қарасақ та, үлгі тұтарлық кісілік, ұстамдылық, талаптылық пен таланттылық, еңбекқорлық қасиеттерді көреміз. Өздерін жоғары орынға апарған да, біздерді сүйсіндірген де осы қасиеттер болар.

В последние годы профессора А. Жубанова приглашают читать лекции студентам по прикладной лингвистике ведущие вузы Алматы, такие как Казахский национальный университет имени аль-Фараби и Алматинский национальный педагогический университет имени Абая. Благодаря своему трудолюбию и в этом деле он имеет большие успехи. Там он нашел учеников, некоторые из них на стадии защиты кандидатской диссертации, а другие только начинают свой путь в науке. До этого времени писавший свои труды на русском языке А. Жубанов на основе прочитанных лекций поставил цель выпустить учебные пособия по прикладной лингвистике на родном языке. Этому могут послужить доказательством такие учебные пособия, как: "Компьютерлік лингвистикаға кіріспе" (Алматы, 2007) – Введение в компьютерную лингвистику, "Қолданбалы лингвистика: қазақ тілінің статистикасы" (2-изд., Алматы, 2007) – Прикладная лингвистика: статистика казахского языка, "Қолданбалы лингвистика: формалды модельдер" (2-изд., Алматы, 2007) – Прикладная лингвистика: формальные модели.

Для человека не обязательно доказывать, что он выдающийся ученый, если в нем это заложено. Каждому представителю Жубановской династии присущи человечность, сдержанность, требовательность и талантливость, трудолюбие. Эти качества дали возможность достигнуть высшего уровня, и они же заставляют нас уважать их.

Өмірде кездесіп жататын өлім-қаза, ауру-сырқау сияқты ауыр жүктерді көтере білу де адамға үлкен сын. Осының бәрі нар көтерер жүк болса, сол жүкті де Асқар көтеріп келеді. Заман мен солақай саясат сынынан өте білген ғалым-азамат тағдыр сынына да арқасын тосып отыр. "Өмірдің жақсы да, жаман да жақтарын басымнан кешірдім, бірақ оның қиын тұстарының бәрі де әкемнің нақақтан "халық жауы" аталуымен салыстырғанда түк те емес сияқты" – дейді ғалым. "Асылдың сынығы, тұлпардың тұяғы", өзі де сом асыл – мықты ғалым, атаулы азамат, үлгілі ұстаз, филология ғылымдарының докторы, профессор Асқар Құдайберген баласы, халқымыз мақтанышпен қастерлеп атайтын "Жұбановтар әулеті" деген қымбат атты жоғары ұстап, кейінгі ұрпақтарға жеткізе бермек, әлі де том-том еңбек жазып, топ-топ шәкірт дайындай бермек!

Рәбиға СЫЗДЫҚ,
академик;
Ғарифолла ӘНЕС,
филология ғылымдарының кандидаты

В жизни бывает много трудностей, перенести их – это огромное испытание для человека. Немало испытаний выпало и на долю Асеке, которые он сумел мужественно победить. "В жизни видел и хорошее, и плохое, но все они не сравнятся с тем, что отца обвинили ложно "врагом народа", – говорит ученый.

Сын великого ученого Аскар Кудайбергенович сам стал видным лингвистом, образцовым педагогом, высоко держащим имя династии Жубановых, которого чтит наш народ. Ждем от него дальнейшей плодотворной работы в научной и педагогической деятельности.

Рабига СЫЗДЫК,
академик;
Гарифулла АНЕС,
кандидат филологических наук

ҚОЛДАНБАЛЫ ЛИНГВИСТИКА: **ҚАЗАҚ ТІЛІНІҢ СТАТИСТИКАСЫ**



Кітапта қолданбалы лингвистика саласы бойынша қазақ тілін зерттеу нәтижелеріне қысқаша ғылыми шолу жасалып, жиілік сөздіктер мәселесі және олардың тілді зерттеу ісінде пайдаланылуы, автоматты түрдегі жасалу жолдары көрсетілді. Тіл зерттеудегі статистикалық әдістің бастама тұстары сипатталып, негізгі сөз таптарының қазақ мәтініндегі үлестірілуінің ықтималды-статистикалық үлгісін (моделін) құрудың проблемалық мәселелері қарастырылды. Тілдің лексика-морфологиялық құрылымына статистикалық әдісті қолданудың алғышарттары берілген.

• АЛҒЫ СӨЗ

• КІРІСПЕ

• *Бірінші тарау*
ЖИІЛІК СӨЗДІКТЕР

• *Екінші тарау*
СТАТИСТИКАЛЫҚ ЛИНГВИСТИКА

• *Үшінші тарау*
ҚАЗАҚ МӘТІНІН ЫҚТИМАЛДЫ-СТАТИСТИКАЛЫҚ МОДЕЛЬДЕУ

• *Төртінші тарау*
ТІЛДІҢ ЛЕКСИКА-МОРФОЛОГИЯЛЫҚ ҚҰРЫЛЫМЫНА
СТАТИСТИКАЛЫҚ ӘДІСТІ ҚОЛДАНУДЫҢ АЛҒЫШАРТТАРЫ

• ҚОРЫТЫНДЫ

• ҚОСЫМША

• ӘДЕБИЕТ

АЛҒЫ СӨЗ

Оқырмандар назарына ұсынылып отырған бұл еңбек* қолданбалы лингвистиканың бір тармағы болып саналатын статистикалық лингвистиканың кейбір өзекті мәселелерін сөз етеді. Автор аса көрнекті ғалым Бодуэн де Куртенэнің: "Тіл білімінде сандық, математикалық ойлауды жиі қолданып, оны мейлінше нақты ғылымдарға жақындату керек" деген ұстанымын басшылыққа алады [22]. Сондай-ақ Р.М.Фрумкинаның "Статистические методы при заведомо неполной информации об объектах исследования предоставляют возможность делать выводы об этих объектах с заданной точностью и надежностью" [91] – деген пікірімен толық келісе отырып, тілге қатысты көптеген тұжырымға статистикалық ғылыми баға беріп, олардың дұрыс-бұрыстығын тексеруге болатындығына арнайы тоқталады.

Ықтималды-статистикалық заңдылықтар табиғат пен қоғамда кездесетін әр алуан құбылыстардың өзара қарым-қатынастарында көрініс табатыны белгілі. Осындай көрініс "жиілікпен" өлшенеді десек, артық айтқандық болмас. Ал тілдік құбылыстың жиілігі – әмбебап тілдік категория. Бұл жерде тіл зерттеу барысында тек математикалық, дәлірек айтсақ, статистикалық әдіс-тәсіл мен математикалық өрнекті (формуланы) қолдану туралы ғана сөз болып отырған жоқ, сонымен бірге сандық сипат белгілі бір тілдің ішкі табиғатына объективті түрде тән деп түсінген жөн. Шынында да, бүгінгі таңда тіл құрылымындағы сапалық және сандық сипаттар іштей тығыз байланыста болатыны ешқандай күмән туғызбайды. Мәселен, кейбір беделді, ірі ғалымдар өз тұжырымының дұрыстығына толықканды дәлелдер келтіруі мүмкін. Осыдан тіл білімінде не басқа ғылым салаларында бірін-бірі терістейтін теориялар мен болжамдар (гипотезалар), тіпті ғылыми мектептер пайда болады. Міне, сондықтан тілді статистикалық тәсілдермен зерттеу тілші-мамандардың субъективті тұжырымдарын не объективті шындыққа айналдыруға көмектеседі, немесе оның теріс екендігін дәлелдеп береді.

Кітап тілдің сөздік қорын зерттеуді мақсат еткен филолог-студент, филолог-мұғалім, филолог-зерттеуші оқырмандарды ең қарапайым түрдегі статистикалық әдістермен және оның қарапайым есеп-қисап жүргізу жолдарымен таныстыруды көздейді.

Сөздік қор – аса күрделі тілдік нысан, бірақ көптеген теориялық және практикалық мәселелер үшін оның кейбір қасиеттерін, яғни ең негізгі деген сандық сипаттарын білу жеткілікті. Мәселен, кейбір жағдайда белгілі көлемдегі мәтін мен оның сөздігі арасындағы жоғары ықтималдыққа ие қатынастарды білсек те жеткілікті екен. Мәтін көлемінің өзгеруіне қарай зерттеу нысанына алынған тілдік бірлік (сөз не сөз тіркес) жиіліктерінің мәтін мен сөздік бойында таралу (үлестірілу) сипаты қандай, алынған сандық нәтижелерді ана тілін, не шет тілін үйретуде қалай қолдануға болады – деген сауалдарға жауап алуға мүмкіндік береді. Сонымен

---

* Бұл оқу құралына автордың орыс тілінде жарық көрген "Квантитативная структура казахского текста" атты монографиясы негіз болды [2]. Бірақ аталған құрал осы еңбектің тікелей аудармасы емес. Өйткені оқу құралы мен ғылыми монографияның құрылымдары мен мақсаты әртүрлі және 1987–2004 жылдар аралығындағы ғылыми жаңалықтар ескеріліп, оқу құралына тиісті толықтырулар енгізілді.

қатар тілдік ақпаратты тарату, ғылыми ақпараттарды автоматты түрде өңдеп, ұзақ мерзімге сақтау, оларды қажеттігіне қарай іздеп, тауып алу мүмкіндігін анықтау да сөздік қордың статистикасын зерттеу арқылы іске асады.

Қолыңыздағы оқу құралына қолданбалы лингвистика саласының "статистикалық тіл білімі" негізгі нысан болуының өзіндік себептері де бар.

Кезінде Бүкілодақ бойынша "Тіл статистикасы" 1957 жылдан зерттеле бастalса, филология ғылымдарының докторы, математик Қ.Б.Бектаев жетекшілік еткен қазақстандық ғылыми шығармашылық топ зерттеу жұмыстарын 1970 жылдан бастап жүргізе бастағаны ғылыми көпшілікке мәлім. ҚазССР ҒА Тіл білімі институтындағы "Статолингвистикалық зерттеу және автоматтандыру" деп аталатын ғылыми шығармашылық топ қазақ әдебиетіндегі белгілі ақын-жазушылар шығармалары мен басқа да стильдерге қатысты мәтіндер тілін статистикалық әдіспен зерттеді. Содан бері осы сала бойынша едәуір ғылыми еңбек жарық көрді. Қазақстандық статолингвист ғалымдардың жұмысына қысқаша шолу жасаумен қатар біз өз зерттеу тәжірибемізді де оқырмандарға таныстыруды мақсат етеміз.

Тілдегі заңдылық жүйеге ғана тән болғандықтан, статистикалық заңдылықтарды ашуға байланысты қолданылатын әдістер кезінде тілдің өзінше бір бөлек жүйе құратындығы туралы сөз ете бермейміз. Басқаша айтсақ, тіл – біріне-бірі қатыссыз құбылыстар жиынтығы емес, керісінше, заңды түрде ұйымдасқан, көпөлшемді және тәуелсіз сипаттағы ақиқат.

Тілдік бірліктердің (дыбыс, морфема, сөз тіркестері, сөйлем) мәтін ішінде (жазба тіл ағымында) кездесуі кездейсоқ оқиға деп есептесек, онда белгілі бір жүйеден туындайтын тілдің өзі де (яғни мәтін) кездейсоқ процеске қатысты болады. Осының салдарынан зерттеу нысанына алынатын нақты бір мәтін белгілі дәрежеде ақиқат сөйлеу тіліне жақын келетін модельденуші мәтінге айналады. Мұндай жағдайда мәтін қасиеттері тұрақты және элементтері бір-біріне тәуелсіз деген қағидатты ұстану қажет болады.

Лингвистикалық статистикада тілдік заңдылықтарды анықтау барысы тәжірибелік байқау мен сол байқаудың негізінде туындайтын нәтижелерді математикалық жолмен өңдеу арқылы іске асады. Сондықтан лингвостатистикалық зерттеулер жаратылыстану ғылымдарындағыдай табиғи және ғылыми тұрғыда жүргізіледі.

Статистикалық деректерді өңдеуде компьютерлік технологияны пайдалану, тілдік мәселелердің шешімін тез арада табу, математикалық статистика әдістерін басқа ғылымдар тәсілдерімен (ақпараттық теория, математикалық логика және т.б.) ұштастыру қазіргі кездегі отандық және шетелдік тіл білімінің даму бағытына сай келеді.

Түркі мәтіндерін статистикалық тәсілдермен зерттеу, ғалымдардың пайымдауынша, бұл тілдердің ішкі табиғатына, яғни олардың жалғамалық (агглютинативтік) құрылымына негізделеді.

Тіл зерттеу тәжірибесінде сандық зерттеулердің қажеттігін математикалық лингвистика мамандарымен қатар дәстүрлі тіл білімінің аса көрнекті өкілдері де мойындап келеді. Олардың пікірінше, жазба әдебиет пен сөйлеу тіліндегі түрлі сөздердің әр ыңғайдағы қолданылу жиілігін анықтау тәсілі тілдің құрылымды-грамматикалық, кейде стильді-семантикалық айырымдарын ажыратумен қатар стильдерге тән қасиеттерді де анықтауға көмектеседі.

Қорыта айтқанда, барлық грамматикалық категорияларды сандық тәсілмен талдау олардың әдеби тілдегі функционалдық салмағын нақты көрсетуге мүмкіндік береді.

Осы оқу құралының қолжазбасымен танысып, құнды-құнды пікір айтқан әл-Фараби атындағы Қазақ ұлттық университеті жалпы тіл білімі кафедрасының меңгерушісі профессор Э.Д.Сүлейменоваға және осы кафедраның профессор-оқытушы құрамына автор өз ризашылығын білдіреді.

КІРІСПЕ

Тілдің қолданбалы аясы бұрыннан да өзінің кең және жан-жақтылығымен ерекшеленетін. Оның ескіден келе жатқан саласы – жазу (графика), ана тілі мен шет тілін оқыту әдістемесі және тілдің лексикографиялық жүйесі. Осылардан кейін барып – аударма, шифртану (дешифровка), орфография, транслитерация және терминологияны өңдеу. Қолданбалы лингвистиканың дәстүрлі бағыттарының бірі – мемлекеттің тіл саясатына қатынасу:

1) әліпбиді, орфографияны өңдеу, сауатсыздықты жою;
2) мемлекеттік тілді тағайындау;
3) мемлекеттік тілден басқа тілдердің орнын анықтау;
4) ұлттық терминологияны өңдеу, бірізге келтіру және қалыптастыру;
5) қала, көше, алаңдарға ат қою мен оның бұрыннан келе жатқан атын өзгерту, яғни ономастика мәселелері.

Қолданбалы тіл білімінің осындай классикалық бағыт-бағдарын дамыту және мейлінше жетілдіру мәселелерімен бірге XX ғасырдың екінші жартысынан бастап оның бірқатар жаңа бағыттары орын ала бастады. Бұл бағыттар заманға сай қоғамдық, жаратылыстану және техникалық ғылымдардың өзара даму сипатынан туындайтын қолданбалы лингвистиканың тарихи дамуының логикалық жалғасының көрінісі еді.

Негізінен алғанда, адам баласының қызметінің әртүрлі аясындағы жұмыстарын жеңілдету бір ғана проблемалық мәселеге тіреледі. Ол – қоғам өміріндегі ақпарат жұмысын өңдеу. Бұндай ақпарат мәтін түріндегі жазба тілінде не үйреншікті сөйлеу тілінде болуы мүмкін.

Қоғамның практикалық қажеттігін өтеуде, әсіресе, ақпаратты сақтау мен оны тарату істерінің сырын ашуда зерттеуші оны ең алдымен тілдің ішкі табиғатындағы заңдылықтардан іздегені жөн.

Осымен байланысты өзекті мәселенің бірі – оператордың дауысымен басқарылатын станок пен құрал-аспаптарды өндіру. Ол үшін әртүрлі тілдердегі ақпараттарды автоматты түрде өңдеу мен оларды іздестіру, байланыс торабы жұмысын (телефон, радиобайланыс т.б.) жетілдіру, адам баласының сөйлеу мен есту қабілетінің бұзылуына қатысты ауруларды емдеуде дыбыстың фонетикалық деректеріне жүгіну және т.б.

Аталған проблемалық мәселелерді ЭЕМ-ды (компьютерді) кең түрде қолдана отырып, мәтінді (жазба, сөйлеу) автоматты түрде өңдеу қажеттілігі туындайды. Мәселен, әр типті ақпараттық жүйелерді тілдік деректермен қамтамасыз ету; машиналық аударма мәселесі; табиғи тілді түсінетін жүйе құру (жасанды интеллект жүйесіндегі тілдік мәселелер); сөйлеу кезіндегі дыбыстық сигналдарда бар ақпараттарды пайдалану негізінде арнайы жүйе құру т.б.

Соңғы жылдары әртүрлі ақпараттық жүйелерді тілдік деректермен қамтамасыз етумен байланысты терминтану мәселесін бірізділікке түсіріп, тұрақтандыру өзекті

мәселеге айналып отыр. Себебі, әр білім саласының ғылыми және техникалық терминдерді қолдану қажеттігі барынша сұраныс тудыруда. Әрине, мұндай мәселе лексикография саласынан тыс өз шешімін таппайтыны белгілі. Сондықтан қоғам өмірінде ЭЕМ-нің (компьютердің) кең қолдануымен тығыз байланысты автоматты лексикография саласы біртіндеп өз отауын құруда.

Практикалық мәселелерді шешуге байланысты табиғи тілде диалогты іске асыру үшін адам мен ЭЕМ арасында әсерлі қарым-қатынас орнату қажет. Мәселен, оларға жататындар: сұрақ-жауап жүйесін құру, роботтарды басқару жүйесін құру, аталған басқару жүйелерінде дұрыс шешім қабылдау үшін диалогтық процесті іске қосу. Мұндағы ең негізгі мәселе – автоматты құрылғының жазба не дыбыстық мәтінді түсіне білу (тану) проблемасы.

Әрбір тілдің қолданбалы мәселені шешуде өзіне тән ерекшелігі бар. Мұндай есептердің санын, түрін алдын ала білу мүмкін емес. Өмірде олар жыл сайын, ай сайын ауысып, біреуі келіп, екіншісі кетіп жатады. Қолданбалы лингвистикаға қатысты іргелі зерттеулер ғылыми-техникалық, ұйымдастыру-басқарудың ауызша-жазбаша құжаттарының, сөздік түзу мен және т.б. түрлерінің фонетикалық, грамматикалық, семантикалық және статистикалық құрылымын сипаттау мен модельдеу жақтарын қамтиды. Жекелеп айтсақ, оған теориялық және қолданбалы лингвистиканың шекарасында жатқан мәтіндік бірліктердің формальды моделін (үлгісін) құру проблемасын жатқызуға болады.

Теориялық лингвистика, негізінен, тілді қалыптық, жүйелік тұрғыда қарастыратындықтан, қолданбалы тіл білімі тілді әрекет үстінде, яғни оның қарым-қатынас (коммуникация) кезіндегі табиғатын түсінуге тырысады. Көптен бері қолданбалы тіл білімі ғылыми-техникалық пен іскерлік прозадан өзін алшақ ұстағаны аян. Дегенмен, XX ғ. 70-жылдарында ғалымдардың мынадай шешім жасауына тура келді. Көптеген қолданбалы мәселелер таза лингвистикалық емес, тіпті дәлірек айтсақ, ол проблеманы шешу адамның іс-әрекеті мен ойлау процесін, тілдің семантикасы мен оның формальды және семантикалық әдіс-амалын синтездеу (жинақтау) жолдарын модельдеуге тікелей қатысты. Осыдан барып қолданбалы тұрғыдағы зерттеулердің іргелі проблемасы – білімді модельдеу (моделирование знаний) мәселесі анықталды.

Соңғы кезде бұл проблема (білімді модельдеу) бірнеше ғылымдар тоғысына, дәлірек айтқанда – логика, лингвистика, математика, психология, кибернетика салаларының аясына қатысты болып отыр. Көптеген ғылымдардың осы мәселеге қызығушылығының негізгі себебі біреу-ақ, ол – автоматты жасанды интеллект жүйесін құру. Солай бола тұра, "білім" бізге "тіл" арқылы берілетіні анық. "Білім" сөйлеу мәтінінде де (монолог, диалог, әр жағдайдағы реплика) және сонымен бірге жазба мәтіндерде де (ескі жазба ескерткіші, көркем әдебиет, ғылыми-техникалық әдебиет) көрініс табады. Тіл арқылы "білімді" біз ұрпақтан-ұрпаққа жеткіземіз. Сондықтан "тіл" – "білімді" сақтау формасы және оны тарату құралы десек те болады. Біздіңше, ғылым мен теханикада мәтіннен тыс жатқан "білім" жоқ деуге болады, ал ғылыми-техникалық мәтіннің семантикасын модельдеу дегеніміз – ол осы салалардың білім жүйесін модельдеу. Осылайша, біртіндеп келіп, білімді модельдеу проблемасы мәтін мазмұнын (мағынасын) модельдеумен ұштасып жатқанына көз жеткізуге болады. Бұл жерде ең негізгі шешімін табуды қажет ететін мәселе – мәтіннің семантикалық көрінісін (семантическое представление) құру.

Дегенмен, техникалық және іскерлік коммуникация аясындағы нақты қолданбалы мәселені шешу сол саладағы құжаттық мәтіндерінің грамматикалық, лексикалық, семантикалық құрылымының сипатталуына байланысты. Сонымен бірге, аталған салалардың терминологиялық сөздігінің болуына, мәтіннің статистикалық құрылымының зерттелуіне және осындай мәтіндер типтерінің семантикалық көрінісінің толық құрылуына да байланысты екенін айта кету қажет.

Қолданбалы лингвистиканың дамуына, яғни ғылыми жетістіктерге ие болуына тіл білімі теориясының дамуы да өз әсерін тигізбей қоймады. Мысалы, XX ғ. 20–30-жылдары практикалық қажеттіктен туындаған ғылыми-техникалық терминологияны бірізді және тұрақты ету жолдары жаңа лингвистикалық пән – "терминтану" ("терминоведение") саласын өмірге әкеледі. Сол сияқты, Кеңес одағы халықтары тілінің әліпбиі мен жазуын құруға байланысты 30–40-жылдары жүргізілген кең көлемді әрі орасан күрделі, әрі зор тәжірибелік мәні бар жұмыстар тілдерді синхронды сипаттаудың әдістерін жетілдіруге себепші болды (ынталандырды) деуге болады.

Соңғы 20–30 жылда пайда болған қолданбалы лингвистиканың жаңа аспектілерінің бәрі бір ғана ортақ проблемаға тәуелді екенін байқатты, ол – тілдің жазба не сөйлеу түрлерін автоматты өңдеу мәселесі еді. Мұндай проблемамен айналысу тілді талдау мен сипаттаудың жаңа әдіс-тәсілдерінің дамуына және тіл табиғаты мен тіл білімі құрылымына деген жаңаша көзқарастың пайда болуына мүмкіндік туғызды.

XX ғасырдың 50–70-жылдарында тілдік материалдарды автоматты түрде өңдеуге қатысты күрделі проблеманың шешім табуы қолданбалы лингвистиканың әрі қарай дамуына айтарлықтай әсер етті.

Қолданбалы лингвистиканың жаңа аспектілері теориялық тіл біліміне жаңадан ғана ене бастаған математикалық әдістерді, әсіресе, теоретика-жиындық, формальды-логикалық, статистика-ықтималдық әдістердің қолдануын барынша жеделдетті.

Классикалық тіл білімінің, классикалық логиканың, психологияның, семантика мен математиканың қиылысуынан барып тілдегі модельдеу әдісінің өзінше бөлек ғылым ретінде көрініс табуының нәтижесінде қазіргі "құрылымдық лингвистика" деп аталып жүрген ғылыми сала өмірге келгені белгілі. Сол сияқты енді тіл біліміндегі дербес тұрған теориялық бағыт ретінде "тілдің теориялық модельдері" атты жаңа тарау пайда болды.

Лингвистика мен математика салаларының аралығынан туындаған – "математикалық лингвистика" атты жаңа пән өмірге келді. Бұл пәннің XX ғ. 50–60-жылдары қалыптасуы бұрыннан да тіл білімінің барлық ішкі дамуының өзінен де сезіле бастаған болатын. Мәселен, бұл жайт тілдің құрылымына көңіл аударудан, тілдік жүйе ретінде қарастырудан және оның микрожүйелерден тұру күрделілігін жете түсінуден, лингвистикалық нысандар мен олардың атрибуттары аралығындағы қатынастарды танып-білуден байқалып, математикалық лингвистика пәнінің дамуына объективті жағдай туғызды деуге болады.

Атап айтқанда, қазіргі теориялық тіл білімінің дамуына ең көп әсер еткен – құрылымдық лингвистика мен математикалық лингвистика пәндері. Бұндай әсер көбінде грамматиканың синтаксис және семантика салаларында айтарлықтай көрініс тапты.

Құрылымдық синтаксисте мынадай формальды екі синтаксистік модельдер жүйелі түрде жете зерттелді. Олар – тікелей құрастырушылар моделі мен өзара тәуелді модельдер деп аталады. Бұл аталған екі модель де машиналық аударма жасауда,

автоматтанған синтаксистік талдау мен мәтінге жасалатын басқа да автоматты өңдеулерде кең түрде пайдаланылады.

Семантика тілдің барлық деңгейлеріне ортақ болғандықтан, тіл білімінде оны бөлек деңгей ретінде қарастырмайды. Семантикада, біріншіден, лингвистиканың өзіне ғана тән әдістің әсері байқалса, екіншіден, логикалық семантика мен теориялық классификациялау әдістерінің де әсері орын алады. Синтаксистік семантикада сан жағынан көбірек зерттелген жайттар – сөйлемнің семантикалық құрылымы жайлы тұжырымдар (концепциялар).

Қоғамдық ғылымдарда ғылыми зерттеулерді автоматтандыру, ең алдымен, құжаттық және фактографикалық деректердің ауқымды қорын құру барысында қажет болды. Себебі, мұндай автоматтандыру типі ең алдымен қоғамдық ғылымдар мамандарының өз қалауынан туындап отыр. Мәселен, ғылым жолындағы қандай анықтамалар түрі қоғамдық саладағы қызметкерге қажеттірек? Біріншіден, мұндай сұраныс кітапханалардағы библиографиялық істе: белгілі бір мәселеге қатысты барлық отандық не шетелдік әдебиеттерді іздестіріп, тауып алу сияқты мәселелер. Мұндай материалдарға жататындар: кітаптар, мақалалар, конференция тезистері, хроникалық шағын мақалалар және т.б.

Екінші кезеңде, жиналған әдебиеттер ішінен қажетті деректерді жинастыру, оларды қажетті белгілеріне қарай сұрыптау, топтастыру және т.б. жұмыстар орындалуы керек.

Көптеген қоғамдық ғылымдарға қатысты мәселелер, әсіресе, тіл білімінде, этнографияда, антропологияда, тарихта карталық мәліметтерге тікелей қатысты болып келеді. Сондықтан, ЭЕМ-ге ондай деректерді енгізу мен қағаз бетіне шығару өте-мөте өзекті деп саналады.

Қоғамдық ғылымдар бойынша ғылыми-зерттеу жұмысын автоматтандыруды қажет ететін мәселелер:

а) әр тілдер бойынша белгілі тақырыпқа қатысты әдебиеттерді іздеп, табу;

ә) жазба ескерткіштер мен мәтіндік материалдар қоры бойынша таңдама жұмыстарын атқару;

б) табылған материалдар арқылы қажетті амал-әрекеттерді іске асыру;

в) материалды жан-жақты сұрыптау;

г) алдын ала белгілі өлшемдер бойынша реестрлер, каталогтар сипаттамаларын жинақтау;

д) статистикалық, картографикалық, сұрыптау теориясы мен жүйелік талдау әдістерін қолдану;

е) берілістерді (деректерді) сызба, сурет, карта түрінде көрсету.

Сонымен, ғылыми жұмысты автоматтандыру процесінде құрастырылған жүйе негізінде тұтынушы маманға қажетті тарихи деңгейде зерттеу нысанына қатысты барынша толық энциклопедиялық білім және библиографиялық мәліметтер берілуі қажет деп санаймыз.

Жоғарыда сөз болған мәселелерді қорыта келе айтатын болсақ, қолданбалы лингвистика, ең алдымен, кешенді пән. Себебі, ол философия, психология, физиология, математика, логика, әлеуметтану, информатика салаларымен үнемі тығыз қатынаста болады.

Жаңаша мағынадағы қолданбалы лингвистиканы жеке қарастыратын болсақ, ол көптармақты ғылыми сала болып табылады. Ғылыми әдебиеттерде оны "компьютерлік лингвистика", "инженерлік лингвистика", "автоматты лингвистика", "есептеу лингвистикасы", "квантитативті лингвистика", "статистикалық

лингвистика" деген жарыспалы терминдермен атап жүр. Әрине, бұл атауларға сәйкес салалардың өздеріне тән бағыт-бағдары ажыратылады, алайда бәріне тән ортақ мақсат – адам баласының қарым-қатынасы үшін қажетті табиғи тіл қызметінің ең ыңғайлы, ең тиімді жолдарын іздестіру.

Қолданбалы лингвистика салаларының барлық тілдерде, соның ішінде қазақ тілінің жазба түріне қатысты зерттеулерде ең көп қолданыс тапқан тармағы – статистикалық лингвистика. Дәстүрлі әдіспен тілді зерттейтін тілші-ғалымдардың бәрі бірдей статистикалық әдіс-тәсілдерді қолдай бермейді, бірақ өз ізденістерінде осы саланың *көп, аз, мол, жиі, сирек, өнімді, өнімсіз* тәрізді терминдерін барынша пайдаланады және олардың статистикалық ғылым саласына қатыстылығын көп жағдайда аңғара бермейді.

Статистикалық лингвистика тілді зерттеудің тек әдісі не тәсілі болып қана қалмай, қазіргі кездегі тілтануда өзінше бір бөлек ғылыми пән дәрежесіне көтеріліп отыр. Бұл пән тілдік бірліктердің сапалық және сандық (мөлшерлік) мәліметтерін тілдің табиғатына сай қарастыруды жөн санайды.

Статистикалық лингвистика, негізінен, математикалық статистика әдістемесіне сүйенеді. Зерттеу барысында ондай әдістемені қолдану қатынас құралы ретіндегі "тілді" белгілі бір жүйе деп қарастырудан туындайды. Шынында да, "тіл" дегеніміз ақиқат болмыста өмір сүретін көп өлшемді және белгілі бір ішкі заңдылықтардың негізінде топталған (реттелген) табиғи құбылыс деуге болады. Ал тілдік жүйе математикалық статистика заңдылықтарына бағыну үшін ондағы бірліктер тобы қайталанып отыратын және кездейсоқтық сипатта болатын тілдік элементтерден тұруы қажет. Бұл жерде айта кететін жайт, тіл қызметі кезінде (яғни тілдік қарым-қатынас кезінде) дыбыстар, сөздер, сөз тіркестері, сөйлемдер және одан да үлкен бөліктер қайталанбайтын сипатта болса, адамдар арасындағы тілдік қатынас бұзылып, ақпарат таратушы мен қабылдаушы арасында түсінбестік пайда болған болар еді. Сондықтан да тілдік бірліктердің қайталануы олардың тілімізге тән болуынан, олардың қолдану сипатына *жиі, сирек* деген ұғымдардың да тән екендігінен туындайды.

Табиғи ортадағы элементтер арасындағы қайталану сипатының негізінде математикалық статистика мен ықтималдық теория атты ғылым салаларында тілдік бірліктердің статистикалық заңдылықтары қалыптасады. Зерттеуге алынған бірліктің (единица, элемент) ондай заңдылыққа бағыну не бағынбауын айқындау үшін алынатын ортаның аумағы (мәселен мәтін көлемі) барынша мол болғаны жөн. Мысалы, кітаптың бір бетінде кездесетін сөздер мен оның жүз бетінде кездесетін сөздердің қайталану сипаты әртүрлі болатындығына ешкімнің күмәны жоқ. Ал заңдылықты айқындау үшін, кітаптың бір бетіндегі мәліметтен гөрі оның көптеген беттерінен алынған мәліметтердің шамасы шындыққа жақын болатындығы белгілі.

Тілге тән заңдылықтарды статистика тәсілімен анықтау негіздемесі мынада:

1) сандық (мөлшерлік) құбылыс тіл табиғатына әуелден-ақ тән болуы;

2) тіл құрылымындағы сандық және сапалық сипаттардың өзара байланыста болуы;

3) тілдің әртүрлі бірліктері сөйлеу ағымында статистикалық заңдылықтардың ең болмағанда біреуіне бағынуы тиіс деп ұйғару керектігі.

Әдістердің индуктивті, дедуктивті болып ажыратылатынын ескере отырып, негізінен алғанда, статистикалық әдіс индуктивті әдіс арқылы, ал тілдің әртүрлі модельдерін жасау үшін қолданылатын логика-математикалық әдістер – дедуктивті әдіс арқылы жүзеге асады.

Тілдің ықтималды-статистикалық моделін жасауға болатыны белгілі десек, ондай модель, біріншіден, белгілі бір тілдік жүйенің мәтіндері арқылы жасалады, ал екіншіден, тілдің логика-математикалық моделі сол тілдің толық жүйесін айқындай алады.

Сөйлесу арқылы қатынас жасау кезіндегі тілдік процесті ***байланыс өзегі*** (канал связи) не ***байланыс жолы*** деп есептеуге болады. Ал мұндай байланыс жолы арқылы таратылатын ақпарат әріп, дыбыс, морфема және т.б. лингвистикалық бірліктер негізінде іске асады. Бұл жердегі лингвистикалық бірліктер белгілі бір "кодтың" символдары ретінде есептеледі. Байланыс өзегіндегі хабарлаушы мен қабылдаушы, яғни айтушы мен тыңдаушы өзара бірдей "кодты" пайдаланулары шарт болуы қажет.

Тіл зерттеу тәжірибесінде статистиканың араласуы бұрыннан да бар десек, ондай әдіс тек жекелеген зерттеушілердің ізденістерінде ғана кездессе, қазір статистикалық лингвистика жеке ғылыми пәнге, ғылыми бағытқа айналып отыр. Осымен байланысты статистикалық лингвистика саласының қазақ тіліне қатысты өзіне тән алға қойған теориялық және практикалық мақсаты мен міндеті әлі де айқындалуда.

Қай ғылым саласын алсаңыз да онда қарастырылатын мәліметтер дәл және объективті болуы шарт, ал қоғамдық ғылымдар ішіндегі осындай дәлдікті көбірек қажет ететін сала – лингвистика.

Тіл зерттеу тәжірибесіне статистикалық әдіс-тәсілді ең алғаш ұсынған ғалым орыстың көрнекті математигі В.Я.Буняковский 1847 жылдың өзінде-ақ ықтималдық жіктеудің мүмкіншіліктерін аса бір білгірлікпен (көрегендікпен) атап, ол істің сәтті болуы үшін филологтар мен математика мамандарының бірлескен одағының қажеттігін айтқан болатын [86].

Тіл саласындағы кез келген статистикалық зерттеуде мынадай үш мәселенің басын ашып алған жөн:

1) Тілдің қай бірлігін, нені санау (есептеу) керек?
2) Таңдалып алынған бірліктерді неге санау (есептеу) керек?
3) Тілдік бірліктерді қалай санау (есептеу) керек?

Яғни зерттеушінің: ***"нені?", "неге?", "қалай?"*** деген үш сұраққа жауабы әр уақытта дайын болғаны жөн.

Бірінші сұрақтың жауабы зерттеушінің алға қойған мақсатына тікелей байланысты. Ең алдымен есептелуге тиісті тілдік бірлік анықталуы қажет және оның тілдегі сапалық сипаты, яғни ол жайлы тілдік мәліметтер жеткілікті түрде жиналуы керек. Саналуға тиісті әрбір бірліктің толық анықтамасы алдын ала айқын болуы талап етіледі. Мәселен, мәтін бойынан: сөйлем, сөзтіркес, сөзқолданыс, сөзформа, сөз, буын, әріп, тыныс белгілер және "бос аралық" (пробел) сияқты бірліктер ішінен қайсысы таңдалып алынса, солардың айырым белгілері, анықтамасы алдын ала белгілі болуы шарт.

Екінші сұрақтың маңыздылығы бірінші сұрақтан кем емес. Статистикалық деректері белгілі тілдік элементтердің бәріне бірдей лингвистикалық түсініктеме беру мүмкін бе, яғни олар лингвистикалық мағынаға ие ме, жоқ па деген сұрақтар туындайды. Сондықтан ***неге санаймыз*** деген сауалға дұрыс жауап іздеу үшін біз болашақ алынатын сандық мәліметтердің тілдік тұрғыдан түсінігі болу-болмауын алдын ала ескерілуі керек.

Әрине, ***неге санаймыз*** деген сұраққа әр кезде бірдей жауап беру көбінесе алға қойған мақсатқа байланысты, сондықтан әр мәселеге өзіне тән жауап іздестіру қажет болады.

Жоғарыда аталған бірінші және екінші сұрақтарға (***не себепті санаймыз)*** жауап беру қажеттігіне осы оқу құралының "БІРІНШІ ТАРАУЫНДА" арнайы тоқталмақпыз.

Енді үшінші – ***"қалай санау"*** сұрағына жауап беру үшін біз алдымен мектепте өткен арифметиканы еске түсіре отыра, ***ықтималдық теориясы және математикалық статистика*** атты жоғары математика пәндерінен хабардар болуымыз керек. Себебі, тілдегі бөліктердің жазба не сөйлеу тілінде аз немесе көп болып келуі кездейсоқтыққа жатады. Ал кездейсоқтықтың да өзіндік заңдылықтары болатынын ескерсек, ол заңдылықты зерттейтін пән – **"Ықтималдықтар теориясы".** Оқу құралының "ЕКІНШІ ТАРАУЫНДА" аталған теорияның ең қарапайым және ең қажетті-ау деген статистикалық құрал түрлері тілші мамандар үшін арнайы қарастырылады.

Енді осы оқу құралының құрылымдық жағына қысқаша тоқталайық. Қазақстандық статолингвист-ғалымдардың негізгі статистикалық құралы – әр стильден түзілген жиілік сөздіктер материалдары болғандығын ескеріп, оқу құралының бірінші тарауы "Жиілік сөздіктер" деп аталып, жиілік сөздіктерге қатысты мәселелер сөз болды.

Ал екінші тарауда Ресейдің белгілі ғалымы Б.Н.Головиннің "Язык и статистика" атты оқу құралының ізімен филолог-студенттерге ең қажетті деген статистикалық минимум құралдар жайлы мәселе қарапайым түрде қарастырылды.

Кітаптың ҮШІНШІ ТАРАУЫ "Қазақ мәтінін ықтималды-статистикалық модельдеу" деп аталып, онда қазақ тіліндегі жиілік сөздіктер материалдары негізінде "Ципф заңы" тексеріледі және жиі қолданыс табатын негізгі сөз таптарының мәтін бойында таралуы теориялық үлестіру заңдарына бағыну-бағынбауының сынау (бағалау) критерийлері арқылы қарастырылды. Соңғы – ТӨРТІНШІ ТАРАУДА қазақ тілі мәтіндеріндегі сөзтұлғалардың лексика-морфологиялық құрылымын статистикалық әдіспен зерттеу үшін шартты белгі-кодты сәйкестендіру бағдарламасы да осы тарауда тұңғыш рет көрініс тауып отыр. Мәтін бірліктеріне тән белгі-кодтарды сәйкестендіру "ұялы принципке" негізделді. Ал негізгі сөз таптары ретінде қазақ тілінің зат есім, етістік, сын есім сөздері тиісті белгі-кодқа сәйкестік кестелер арқылы берілді.

Оқу құралында қолданыс тапқан пәндік терминдердің көбіне автордың өз қалауымен алынғанына байланысты мемлекеттік деңгейде бекіп үлгермеген кейбір атаулардың кездесуі мүмкін екендігін ескерте кеткіміз келеді. Осы кітапта көтерілген мәселелер жайлы жайттарды, негізінен, орысша жазылған әдебиеттерден кездестіруге болатындықтан, кітаптың соңында "Қосымша" ретінде оқу құралының мәтінінде кездесетін терминдер мен олардың фразалық тіркестерінің қазақша-орысша сөздігі және басқа да қолданбалы лингвистика мен математикалық лингвистика пәндерінде кездесетін негізгі терминдерінің кейбір атауларының қысқаша орысша-қазақша сөздігі мен анықтамалары берілді.

Бірінші тарау
ЖИЛІК СӨЗДІКТЕР

1.1. Қазақ тілтанымындағы статистикалық әдістің орны

Ғылыми-техникалық прогресс, электронды-есептеу техникасының дамуы, нақты әдіс-тәсілдердің түрлі білімдер салаларында қолданыс табуы тіл біліміне, соның ішінде түркітануға да өз ықпалын тигізді.

Өткен ғасырдың өзінде үндіеуропа тіл білімінде кең тараған статистикалық әдіс түркітануда, соның ішінде қазақ тілінде тек XX ғасырдың 60-жылдарынан бастап жүйелі және ғылыми тұрғыда қолданыла бастады. Бірақ бұдан бұрын да аталған әдіс ішінара қолданыс тауып, тілдің әр деңгейіндегі бірліктердің сандық және пайыздық арасалмағын анықтаумен шектеліп келген болатын.

Түркі тіліндегі мәтіндерді статистикалық тәсілмен зерттеу – бұл тілдің ішкі табиғатына, яғни агглютинативті құрылымына негізделеді. Мамандардың пайымдауынша, кітаби және сөйлеу тіліндегі түрлі типті сөздердің әр ыңғайдағы қолдану жиілігін анықтайтын мұндай тәсіл тілдің құрылымдық-грамматикалық, кейде стильдік-семантикалық айырым белгілерін ажырата отырып, осы стильдерге тән қасиеттерді анықтауға мүмкіндік жасайды. Сонымен бірге грамматикалық категорияларды сандық тәсілмен талдау әдеби тілдің барлық аясындағы функционалдық салмағын көрсетуде де маңызды.

Белгілі филолог-статист Р.Г.Пиотровский түркі тілдерін сандық деңгейде зерттеудің негізі олардың ішкі құрылымдық табиғатына тән екендігін айта келе, бұл тілдерге үндіеуропа тіліне тән екпінді ассимиляция мен редукциялар және осы сипатты қасиеттер жат екенін баса көрсетеді. Бұл жағдай тілдік деректердің жай кезде қиындық туғызатын кейбір тұстарын математикалық әдіспен тіркеу мен тануда аса маңызды деп саналады [75].

Тілдік мәселелерге қолданылатын математикалық әдісті интуитивті түрде тұжырымдалған, сол сияқты толық шешімі жоқ мәселелердің логикалық тұжырымдалуына бағынатын және алгоритмдік шешімі бар бір немесе бірнеше қарапайым математикалық есептермен ауыстыру деп ұғыну қажет. Тілдік мәселені осылайша математикалық жолмен шешу – лингвистикалық нысанның математикалық моделіне (үлгісіне) көшуді қажет етеді [31, 6]. Тілдің аса күрделі түрдегі сан алуан параметрлі көп қабатты жүйе екенін ескерсек, онда ықтималдық теория, математикалық статистика, ақпарат теориясы және т.б. ғылыми салалар әдістерін қолданудың мүмкіндігі кең екені байқалады. Әсіресе, бұл әдістер сөйлеу тәжірибесінде (жазба не ауызша) жиі кездесетін белгілі бір кешенді шарттарға байланысты тілдік кездейсоқтық құбылыс заңдылықтарын қарастырғанда аса ыңғайлы. Тіл табиғатының нақты ғылымдар саласының ережелеріне бағына бермейтін тұстары баршылық. Сондықтан осындай күрделі нысан ретіндегі тіл саласына ықтималдық теориясы мен математикалық статистиканы қолдану үшін, ең алдымен, "тілге" бірнеше шектеулер қою қажет. Мәселен, толығымен өзгермелі бірліктерден тұратын табиғатына жүйелілік сипат тән тілді көптеген зерттеуші

"ашық жүйе" ретінде қарастыруды дұрыс деп санайды. Осы тұжырымның негізінде, тілге қойылатын бірінші шектеу бойынша, белгілі бір кезеңде тілдік жүйеде өзгермелілік сипат болмайды деп, оны "жабық жүйе" ретінде санау (модельдеу) ұйғарылды.

Екінші шектеу – тілдің ауызша және жазбаша түрлерінің аралық айырым белгілерін ескермей, оны арнайы түрде құрастырылған белгілер жүйесі немесе жазба тіл деп қарастыру. Мұндай ұйғарым мәтінді көру арқылы қабылдауға негізделген сипаттау грамматикасын құруға жеткілікті негіз бола алады [74].

Статистикалық зерттеу кезінде нақты мәтінді сөйлеу ақиқаттығына белгілі дәрежеде сәйкес келеді деп ұйғару негізінде оның үлгісі зерттеу нысаны ретінде алынады. Әрине, бұл жағдайда мәтін қасиеттері арнайы кеңейтіліп қарастырылады және мәтін үлгісі бір-біріне тәуелсіз дискретті бірліктердің тізбегі ретінде танылады.

Тілге қойылатын келесі шектеудің туындау себебі: тілдің қай саласында болмасын зерттеу арқылы сөйлеу қызметінің әр алуандығын қамту мүмкін еместігінде. Өйткені зерттеу нысаны ретінде алынған мәтінге қатысты мәліметтер тек сол мәтіндік стильге ғана тән болады. Зерттеу барысында мәтін типі мен оның сандық мөлшерін дұрыс таңдай білу зерттеу арқылы байқалған заңдылықтың дұрыстығына күмән туғызбаудың айғағы деп саналады.

Статистикалық әдіс басқа математикалық әдістерге қарағанда тіл зерттеу ісіне бұрынырақ және тұрақты сипатта енді деуге болады. Бұл әдістің қажеттілігі айқын, ол қазіргі жағдайда тілшілер арасында айтарлықтай күмән туғызбайды.

Белгілі орыс ғалымы Б.Н.Головин әуелден-ақ сандық сипаттың тіл категориялары мен оның элементтеріне объективті түрде тән болуынан, сонымен бірге сапалық және сандық сипаттарының өзара қатынастылығы тілдің қатынас құралы ретіндегі қызметі мен даму сатысындағы объективтілігінен тілді сандық тұрғыда зерттеу қажеттілігі туындады деп түсіндіреді [30]. Мысалы, морфемдік құрам сөздің сапасына әсер ететіндігі сияқты морфологиялық деңгейдегі сандық құбылыстар оның синтаксистік деңгейдегі сапасына әсер етпей қоймайды. Осыдан барып әр қалам иесінің стилі де тек өзіне ғана тән тұрақты бірліктерінің қатынастары арқылы ерекшеленеді.

Заңдылықтарды тәжірибелік байқау мен оның нәтижелерін математикалық жолмен өңдеу арқылы анықтау тілдік жүйені танып-білудегі лингвистикалық статистикадағы әрі табиғи, әрі ғылыми бетбұрыс болып саналады. Нәтиженің дәлдігіне, жасалатын тұжырымдардың объективті болуына талпыну – барлық ғылымдар саласына тән жағдай. Ал қоғамдық ғылымдар саласындағы ерекше дәлдікті қажет ететін ғылым саласы – лингвистика деуге болады.

Статистикалық әдістің тіл саласындағы қолданылу тарихын сөз еткенде, осы әдісті тұңғыш ұсынған орыстың белгілі математик-ғалымы В.Я. Буняковскийдің (1804–1889) атын атамай кетуге болмайды [86]. Сонымен бірге басқа да ірі ғалымдардың тілді зерттеуде сандық деректерге жүгінудің қажеттілігін дәлелдеулеріне деректерді көптеп келтіруге болады (И.А. Бодуэн де Куртенэ, А.М. Пешковский, М.Н. Петерсон, Е.Д. Поливанов, В.В. Виноградов және т.б.).

Классикалық тіл білімінің өкілі В.В. Виноградов тілдің кітаби, сөйлеу түрлерінде және көркем әдебиет стилінің әр жанрларында сөздердің қолдану жиілігінің біркелкі еместігін айта келе, оларды анықтау үшін арнайы зерттеудің қажеттігін көрсетеді. Ғалымның тұжырымдауынша, мұндай зерттеулер түрлі стильдердің құрылымдық-грамматикалық, тіпті семантикалық айырмашылықтарын белгілеуге, сондай-ақ барлық грамматикалық категорияларына

статистикалық талдау жүргізілген әр әдеби стильдің қатынастық-функционалдық салмағын айқындауға мүмкіндік туғызады [26].

Мәселен, Ф.П.Филин өзінің бір еңбегінде: "Ғылымның бір-ақ мақсаты болады, ол – әлі ашылмаған шындықты ашу немесе соның ашылуына ықпал жасау, сол арқылы қоғамға қызмет ету. Алға қойған мұндай мақсатқа әртүрлі жолмен жетуге болады. Соның ішінде зерттеудің дәлдігі мен объективтілігін қамтамасыз ететін тәсіл ғана ең дұрысы болмақ. Дәлдік пен объективтілікке ұмтылу – қазіргі ғылымның туы. Бұл тіл біліміне де қатысты", – деп жазса [89], В.Н.Головин тіл ғылымына сандық сипаттың тән екендігін, сондықтан тілші-ғалымдардың көпшілігінің тілдік деректерді сипаттағанда сандық ұғымдарға жүгінетінін айтады [30].

Тілдік фактілерді сандық ұғымдарға жүгініп баяндау – қазақ тіл білімін зерттеуші ғалымдарға да тән. Өйткені қай кездегі болмасын қазақ тілі жайлы зерттеулердің бәрінде дерлік ***аз, көп, көптеген, жиі, сирек, өнімді, өнімсіз, мол, жуық, тең, шамалас, көп рет*** т.б. шама, мөлшерді көрсететін сөздер кездесіп отырады және бұлардың мәнін нақты санмен көрсетуге талаптанушы ғалымдар да бар. Сөйтіп, қолданбалы тіл білімінің бүгінде "Статистикалық лингвистика" деп аталып жүрген жаңа саласының алғашқы нышандары мен белгілері міне осылардан басталады. Ал қазақ тіл біліміндегі бұл саланың бастауында профессор Құдайберген Қуанұлы Жұбанов тұрды. Мұны белгілі тілші-статист ғалымдар – Қ.Б.Бектаев, С.Мырзабеков т.б. қазақ тілінің статистикасы туралы зерттеулері мен мақалаларында атап көрсеткен. Мәселен, С.Мырзабеков "Қазақ тілін зерттеуде санды пайдалану" атты мақаласында: "Қазақ совет тіл білімінің негізін қалаушылардың бірі, совет дәуірінде қазақ тілін тұңғыш рет ғылыми түрде зерттеген маман-лингвист Қ.Жұбанов ретіне қарай сандық деректерді де пайдаланған", – десе [69, 167], ғалым Қ.Б.Бектаев "Лингвистикалық статистикаға қатысты элементтерді – сандық деректер мен мәліметтерді, оны пайдалану нәтижесіндегі кейбір пікірлер мен тұжырымдарды біз профессор Қ.Жұбановтың еңбектерінен жиі кездестіреміз", – дейді [10, 110]. Ғалымның сол кездің өзінде-ақ тіл білімінде сандық деректерді пайдаланудың қажеттілігіне ерекше ден қойғаны байқалады. Мәселен, ғалымның 1936 жылы жарық көрген "Қазақ тілінің грамматикасында" сөз мүшелерінің орны, дыбыстардың саны мен сапасы, түрлері туралы жазған мына тұжырымы көңілге қонымды: "Дыбыстың саны бар да, сапасы бар. Дыбыстың сапасы бірдей бола тұрып, саны әлденешеу бола береді. ...Бір тілде жұмсалатын дыбыстардың жалпы санының ұшы-қиыры жоқ, бірақ оның бәрінің бірдей сапасы әртүрлі бола бермейді. Сапасы әртүрлі болатын дыбыстар да көп болмайды". Сол сияқты "... Ілиястың небәрі 12 жол "Қойшы ойы" деген кішкене өлеңінде 488 дыбыс бар. Бірақ 488 дыбыс түрлі емес, мұнда небәры 28-ақ түрлі дыбыс бар. Осы 28 түрлі дыбыстың бірі 10; бірі 50 рет қайталанып барып 488 болған. ... Қазақ тілінде де сапасы әртүрлі болатын дыбыстардың арнаулы саны бар" [40, 183].

Профессор Қ.Қ.Жұбановтың "Қазақ тілі жөніндегі зерттеулерінде" бұл жайлардан басқа да мәселелер, айталық: сөз тіркесінің формалары, сөйлем мүшелерінің орын тәртібі, буындар және ондағы дыбыс саны, буындарды оқыту әдістемесі т.б. мәселелер сөз болған тұста көптеген сандық (статистикалық) деректер мен мәліметтер келтірілген [40].

Шын мәніндегі статистикалық ізденістерді Қ.Жұбанов қазақ орфографиясын ғылыми негізде құру мақсатында қолданған. Ол әліпби құрамына енуге тиісті әріптердің санын қазақ тіліндегі белгілі бір фонеманы іс жүзіндегі қажеттігіне,

оған деген ділгерлікке, қолданыс жиілігіне бағындыра белгілеуді мақсат тұтқан [10, 114].

Міне, бұл айтылғандардан проф. Қ.Қ.Жұбановтың тілді зерттеуде, оның заңдылықтарын қарауда есеп-санаққа, ретті жерінде математикалық ойлауға, оның тәсілдеріне сүйенгендігінің тағы да бір дәлелі.

Қ.Қ.Жұбанов еңбектеріндегі сандық деректерді пайдалану жайымен таныса отырып, мынадай қорытынды жасауға болады: проф. Құдайберген Қуанұлы Жұбановтың қазақ тіл білімі туралы зерттеулерінде тілді сандық деректер мен мәліметтер негізінде қараудың алғашқы нышандары бар. Бүгінде тіл білімінің бір саласына айналып отырған лингвистикалық статистиканы ғалым сол кездің өзінде-ақ тани білген. Тілдің сандық сипатынан байқалатын сапалық қасиеттерін оны зерттеуде пайдалануға болатын тиімді әдіс көздерінің бірі екенін де байқаған. Сөйтіп, ол сандық деректердің негізінде тілге қатысты бірсыпыра жайларды (дыбыстар, буындар, сөздер, сөйлем мүшелерінің орны т.б.) нақты анықтап, кейбір тың тұжырымдар жасаған. Бұлардың маңыздылығы сонда – олар қазіргі кезде бұл саладағы ізденуші-зерттеушілерге үлгі-тірек, бағыт-бағдар болып отыр [10, 115–116].

Қазақ тілінің нормативті курсын жүйелеу жұмысына проф. Қ.Жұбановтан кейін белгілі тілші-ғалымдар: Н.Т.Сауранбаев, І.К.Кеңесбаев, М.Б.Балақаев, А.Ы.Ысқақов және т.б. ғалымдар атсалысты. Бастауыш және жоғары сыныптарға және педучилищелерге арналған қазақ тілі оқулықтарының авторлары тілдің буындық құрылымын, дыбыс физиологиясын, тілдің акценттік және екпіндік және т.б. мәселелеріне қатысты зерттеулерінде сан мен сапаны бөлмей бірге қарастырады. Мәселен, І.К.Кеңесбаев қазақ тіліндегі буын типтерінің, М.Б.Балақаев қазақ тіліндегі синтаксистік сөз тіркестерінің табиғатын зерттеуде көптеген нақты мәтіндік материалдарды сандық қатынастар көмегімен қарастырған [8, 52]. Арнайы сандық талдауларды профессор Ғ.Мұсабаевтың зерттеулерінен де кездестіруге болады [68]. Әрине, ғалымдардың мұндай ізденістері, яғни сандық қатынасқа көңіл бөлуі саналы түрдегі ықыластан гөрі тілдің ішкі табиғатына (сырына) қатысты туындаған деп пайымдауға болады.

Осы кезеңде қазақ тілінің грамматикалық форманттарын статистикалық жолмен зерттеуге айрықша көңіл бөліне бастаған болатын. Қазақ лексикографтары сандық деректі ұлы ақын Абай тілінің сөздігін түзу кезінде жүйелі түрде толық пайдаланды. Тұңғыш рет қазақ тіл білімінің тарихында бір ғана автордың (Абайдың) жазба түріндегі шығармалар тілі толық қамтылып, әр сөзі мен сөз тіркесінің қолдану жиілігіне көңіл бөлінді. Яғни Абайдың тіл байлығы, саны мен сапасы жайлы жайттар тілшілердің арнайы зерттеу нысанына айналды. Сөйтіп, орыстың ұлы ақыны Пушкин тілі сөздігі тәріздес – "Абай тілі сөздігі" жасалды [1]. Бұл сөздікті әрі түсіндірме сөздік, әрі әліпби-жиілік сөздік деуге болады. Сөздердің қолдану жиілігін көрсету сөздік түзудің негізгі мақсаттарының бірі болып саналды. Шығармадағы сөздердің жиілігін ескеру дәстүрлі картотека жасау жолымен іске асты.

1969 жылы Қазақстан Ғылым академиясы Тіл білімі институтында "Статистическое и информационное изучение тюркских языков" атты Бүкілодақтық ғылыми жиынның өтуі қазақ тіл білімі үшін елеулі құбылыс болды. Осы ғылыми форумның нәтижесі ретінде 1970 жылы Тіл білімі институтында статолингвистика саласының белгілі ғалымы Қалдыбай Бектаевтың жетекшілік етуімен "Тіл статистикасы және автоматтандыру" атты ғылыми топ құрылды. Осы кезеңнен

бастап қазақ тілін статистикалық зерттеу ісі жүйелі сипатқа ауысты. Бірнеше жас ғалым статистикалық лингвистика саласынан кандидаттық диссертация қорғап, олардың ғылыми мақалалар жинағы, монографиялары, оқу құралдары жарық көрді. Кеңес одағы кеңістігіндегі ғылыми конференцияларға кеңінен жол ашылды. Қазақстан статолингвистері зерттеулерінің нәтижесі ретінде 1973 жылы "Қазақ тексінің статистикасы" атты ғылыми жинақ жарық көрді [58]. Бұл жинақ сол жылға дейінгі қорғалған (және дайындалған) кандидаттық диссертациялардың негізгі мазмұнын ашады деуге болады.

1973–1974 жж. Р.Г.Пиотровский мен Қ.Б.Бектаевтың "Математические методы в языкознании" атты екі бөлімнен тұратын, жоғары оқу орындарына арналған оқу құралының Қазақ мемлекеттік университетінен жарық көруі де осы жаңа саланың жетістігі еді [11].

Тілді зерттеу тәжірибесінде әр тілді жалпы типологиялық тұрғыдан бағалауда қолданылатын статистикалық критерийлердің үлес салмағын "өлшеуге" (анықтауға) қажетті жалпылама сандық дәреже әлі де ғалымдарды толық қанағаттандырмай келеді. Міне, осы бос кеңістікті толтыруға арналған іргелі ғылыми зерттеу ретінде 1978 жылы басылып шыққан Қ.Б.Бектаевтың "Статистико-информационная типология тюркского текста" атты монографиясын атауға болады [12]. Бұл зерттеу түркі мәтіндері үшін ықтималдық және ақпараттық үлгілеу әдістерін жүйелі түрде қолдануға арналды. Ғалым түркі тілдерінің типологиялық ерекшеліктерін үндіеуропа тілдерімен салыстыра зерттеуді мақсат етіп қойып, оны формальды статистикалық-ықтималдық әдіс арқылы іске асырды.

Осы кезеңдік зерттеулерге тән мәселелер – ол тілдің фонетика мен фонология саласы, сөз бен сөзтұлғаның буындық құрылымы, қазақ мәтінінің түрлі функционалды стилінің лексика-морфологиялық құрылымы, қазақ тілінің ақпараттық сипаты және т.б. кешенді проблемалар болатын. Негізгі зерттеуге қажетті статистикалық құрал ретінде электронды-есептеу машиналары (ЭЕМ) көмегімен жасалатын әліпби-жиілік сөздік, жиілік сөздік және кері әліпби жиілік сөздіктер материалдары болды. Жиілік сөздіктерді ЭЕМ көмегімен автоматты түрде түзу мәселесі мен олардың құнды деректерін тілді зерттеу практикасында іске асыру мүмкіндігі жайлы Кеңес одағы кеңістігі ғылыми орталықтарында ұйымдастырылған көптеген Бүкілодақтық ғылыми жиындар болып өтті. Қазақ тілін статистикалық әдіспен зерттеу және оны автоматтандыру мәселелері де одақтық дәрежеде қарастырылды. Отандық түркітану ғылымы үшін ортағасырлық жазба ескерткіші "Кодекс Куманикус" мәтіні тұңғыш рет ЭЕМ жадына енгізіліп, ол бойынша автоматты түрде "Куманша-қазақша жиілік сөздігінің" алынуы қазақ тілі тарихы үшін елеулі оқиға деуге болады [61].

Жоғарыда сөз болған қазақ тілі мәтіндерін статистика тәсілімен зерттеуге арналған кандидаттық диссертациялар тақырыбы 1979 жылы "Ғылым" баспасынан жарық көрген М.О.Әуезовтің "Абай жолы" романының жиілік сөздіктеріне тікелей не салыстырмалы тұрғыда қатысты болды [13]. Ал ЭЕМ көмегімен алынған және 1995 жылы басылып шыққан "М.Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздіктері" ұлы жазушының 100 жылдық мерейтойына арналған қазақстандық статолингвистерінің арнайы сыйы деуге болады [14]. Жазушы тілінің лексикалық байлығын, автордың сөз қолданудағы стильдік тәсілінің өзіндік ерекшелігін, қазақ тілінің көркемдік және бейнелеуші мүмкіндіктерін меңгеру дәрежесін осы аталған лексикографиялық еңбек бойынша бағалауға болады. Осымен қатар бұл сөздік жалпы қазақ лексикасын статистикалық құрылым тұрғысынан зерттеуге бай тілдік

материал бола алатыны сөзсіз. Кітап түрінде баспадан жарық көрмегенімен М.Әуезовтің әр томы бойынша компьютер жадында сақталған "Сөзнұсқағыш әліпби жиілік сөздік" атты электронды сөздік базасы да арнайы қаралатын лексикографиялық еңбек деуге тұрады.

А.Ахабаевтың қазақ тілін нормативті тұрғыда тануға, қазіргі қазақ публицистикасы тілінің лексика-морфологиялық құрылымын статистикалық талдауға арналған ғылыми жұмысы да маңызды зерттеу болып табылмақ [3]. Бұл жұмыстың шағын бөлігі дублет сөздер вариантының нормасын статистика жолымен айқындау мәселесіне арналған. Статистикалық талдау негізінде А.Ахабаев сөздердің жиілік сипаты негізгі объективті критерий болатындығын сөз етеді. Жалпы алғанда, бұл ғылыми жұмыс 1965–1966 жж. газет мәтіндері мен "Абай жолы" романы (2-ші кітап) мәтініндегі зат есім мен есімдік жалғауларының форма-вариантарын статистика-морфологиялық тұрғыда салыстыра талдауға арналып, соның негізінде қазақ тілінің грамматикалық тұлғаларының нұсқалары стильдік айырым белгілер болатыны анықталған.

Статистика-морфологиялық зерттеудің келесі бір нысаны – қазіргі қазақ тілі етістік сөздерінің туынды тұлғаларын құрайтын негіз етістік пен оларға жалғанатын аффикстер. Осы зерттеуді М.Әуезовтің "Абай жолы" романындағы туынды түбір етістіктерінің құрылымына статистикалық талдау жасау арқылы ғалым С.Мырзабеков жүргізді. Ғалымның сандық деректері бойынша романда кездесетін барлық түбір етістіктің 80 пайызы – туынды етістіктер екені және етістіктің туынды түбірлері негізгілерден төрт есе көп қолданылатыны анықталды [70]. Етістік құрылымына жасалған лингва-статистикалық талдаудың нәтижелері қазіргі қазақ тіліндегі етістік сөздерді танып-білуге мүмкіндік беретіні сөзсіз.

Қазақ тіліндегі жалғаулар мен олардың құрылымдық тұлғалары, терминдер мен кірме сөздер, сөз тіркестері статистикасы сияқты мәселелер А.Белботаевтың ғылыми ізденістерінде қарастырылды [18].

Тіл ерекшелігі мәтіндердің жанрлық (стильдік) түрлеріне қарай да айырым табатыны белгілі. Осындай айырымдарды қазақ тілінің әртүрлі жанрларына қатысты статистикалық жолмен тағайындауды қазақстандық статистика тобының ғалымдары 70–80-жылдары өз мақсаты етіп қойды. Мысалы, Қ.Молдабеков пен Б.Е.Қалыбековтер балалар әдебиеті мен бастауыш сыныптарының оқулықтарын зерттеуді өз міндеттеріне алса, А.Р.Зекенова – М.Әуезовтің драмалық шығармалар тілін, А.Ахабаев – газет тілін, А.Белботаев – ғылыми-техникалық стиль, математикалық стиль тілдерін өз зерттеу нысандары етіп алып, алғашқы ғылыми нәтижелерге қол жеткізді [3, 18, 19, 43, 60, 67].

Қазақ тілінің фонетика саласы да статистика әдісінің назарынан тыс қалған жоқ. Фонетиканың дәстүрлі (есту) және құралды-тәжірибелік зерттеу әдісі бойынша фонологиялық және графикалық жүйелерге қатысты мәселелерін ықтималды-статистикалық әдіс-тәсілдермен шешу тиімді деп танылды [9, 32, 33, 53, 87].

Қазақ тіл білімінің қолданбалы саласының тың проблемалық мәселесі – қазақ мәтінінің әр әріпке (дыбысқа) шаққандағы ақпараттық өлшемін, яғни ***энтропияны*** анықтау. Осындай зерттеулер белгілі ғалым Қ.Б.Бектаев пен оның шәкірті Д.А.Байтанаеваның қазақ мәтініне қатысты ақпараттық-статистикалық ғылыми ізденістері арқылы іске асырыла бастады. Табиғи тілдің белгілер жүйесі үш түрлі деңгейде – ***синтактикалық, семантикалық*** және ***прагматикалық*** аспектілерде қарастырылатыны мәлім. Мұндағы синтактика деңгейі әртүрлі белгілер жүйесінің синтаксисін, яғни сөз мағынасынан тыс, белгілердің тіркесімдік құрылымын және

олардың туындау ережелерімен қатар белгілер құрылымының ішкі қасиеттерін қарастырады. Ал белгілер жүйесінің сыртқы құрылымдық қасиеттерімен семантика мен прагматика айналысады. Дәлірек айтсақ, егер ***семантика*** белгілер жүйесін мағыналық көзқарас, яғни мәтін мазмұны тұрғысынан танып-білумен шұғылданса, ***прагматика*** – белгілер жүйесінің қабылдаушыға қатынасын анықтау мәселелерімен айналысады.

Қазіргі кезде ақпарат аясын теориялық және технологиялық тұрғыда зерттеудің синтактикалық деңгейі жақсы дамып келеді. Бұған К.Шенноннның символдар жүйелілігі мен комбинаторикасын қарастыратын статистикалық байланыс теориясы негіз болады. Бұл теория бойынша табиғи тілде жазылған мәтін белгілі бір код құрайтын дискретті белгілердің тізбегі ретінде қарастырылады. Ал мұндай мәтінді қабылдау кезінде дискретті белгілердің қай-қайсысының да пайда болуы белгісіздіктің шамасымен, яғни энтропиямен өлшенеді. Қ.Б.Бектаев пен Д.А.Байтанаеваның ғылыми ізденістерінде қазақ мәтінінің ең кіші бірлігі – әріпке түсетін энтропияның орта шамасын және қазақ тілінің толық және стильдік (жанрлық) айырымдарын анықтайтын жазба мәтіндеріндегі энтропия мен артық ақпарат шамаларын анықтау негізгі мақсат ретінде қойылды. Осымен қатар қазақ тілі бойынша анықталған осындай сипаттамаларды синтактикалық деңгейде үндіеуропа тілдері бойынша алынған белгілі деректермен салыстыра зерттеулер де іске асты. Қазақ мәтініндегі әрбір әріпке түсетін энтропияның ең үлкен (максимум) шамасын анықтау, әліпби әріптерін қабылдаудың ықтималдықтары тең және өзара тәуелсіз дәрежеде деген ұйғарымға негізделетіні айқын [5]. Осы аталғандармен бірге Д.А.Байтанаева қазақ жазба мәтінінің буын бірлігінің статистика-ақпараттық қасиеттерін тәжірибелік мәліметтер негізінде айқындайтын статистикалық зерттеулер де жүргізді [6].

Кезінде Бүкілодақтық статистика тобы мүшелігінде Қазақстаннан басқа да түркі тілдес мемлекеттердің статист-тілшілері болды. Осының нәтижесінде 1988 жылы К.Дыйқановтың қырғыздың "Манас" эпосының 2 бөлімнен тұратын жиілік сөздігі жарық көрді. Ж.Жетешиков қырғыз публицистика тіліндегі зат есімдердің сөз өзгертуші аффикстерінің статистикалық деректерін қазақ тілінің осындай мәліметтерімен салыстыра зерттеді [38]. С.И.Ибрагимов ғылыми-техникалық стильдері мәтіндері бойынша қырғыз тілінің зат есімдерінің синтаксисті-функционалды кластарын бөлек қарастырып, олардың басқа сөз таптарымен синтаксистік байланыс түрлерін статистикалық жолмен қарастырды [48].

Қырғыз тілі бойынша жүргізілген іргелі ғылыми-зерттеу жұмыстары қатарына Т.Садықовтың "Проблемы моделирования тюркской морфологии" (аспект порождения киргизской именной словоформы) атты монографиясын [84] қосуға болады. Бұл жұмыста қырғыз тілінің алдын ала берілген лексика-грамматикалық ақпараты бойынша есім сөзтұлғалардың туындау үлгісін (моделін) құру көзделді. Ол үшін қырғыз тілінің морфологиялық құрылымына жүйелі түрде талдау жасалды [84].

Қазіргі өзбек тілінің статистикалық құрылымын зерттеушілердің бірі С.А.Ризаев болды. Ол өзбек әдеби тіліндегі екі фонемдік тіркесім жағдайын қарастырып, олардың қатынастық жиілікке әсер ететін факторларын анықтады. Публицистика, ғылыми және көркем әдебиет (балалар ертегісі) стильдерінің түрлі жанрларындағы фонеманың сандық қатынастарын қазақ тіліндегі осындай деректермен статистика жолымен салыстыра зерттеп, олардың өзіне тән ерекшеліктері негізінде фонемалық топтарын бөліп алуға мүмкіндік туды.

Қазақ тілі мен өзбек тілдерінің фонемалық спектрлерін статистикалық талдауға А.Жүнісбеков пен Д.Маматов өз зерттеулерін арнады [34, 64]. Бұл жұмыстарда

қазақ тілінің дауысты, ал өзбек тіліндегі күрделі дауыссыз дыбыстардың спектрлері қарастырылды. Осының негізінде дербес фонемалар мен олардың позициялық вариантары жайлы статистикалық деректер алынды. Осы тәріздес зерттеу жұмыстары қазақ, қарақалпақ пен әзірбайжан тілдері бойынша да жүргізілді [7, 23, 72].

С.А.Ризаев өзбек тілінің буын типтерін және олардың сөздердегі позициялық орналасуы мен ашық, жабық буындардың қатынастық сипаттары негізінде сөздердің тұрақты (канондық) тұрпатын анықтады. Мұндай зерттеулер балалар әдебиеті бойынша да жүргізілді [79].

Қазіргі татар әдеби тіліндегі буын саны мен олардың буындық құрылым типтері, комбинаторлық қасиеті, сөздердегі буындардың қолдану жиілігі жайлы мәселелерді Т.И.Ибрагимовтың ғылыми ізденістерінен кездестіруге болады [49]. Мәселен, татар тілі сөздеріндегі алдыңғы буын өзінен кейін келетін буын жайлы ақпараттың 47 пайызын өз бойында ұстайтындығы анықталды. Мұндай деректер тілдің құрылымдық және фонетикалық құрылымы жайлы көптеген мәлімет беретіні белгілі. Буынның ішкі құрылымына және сөздің буындық құрылымына ықтималды-статистикалық әдіс қолдану арқылы зерттеу өзбек және татар тілдерінен басқа қазақ тілі бойынша да молынан жүргізілді деуге болады. Мәселен, Қ.Б.Бектаев көркем әдебиет, публицистика, ғылыми-техникалық мәтіндер мен 2 томдық "Қазақ тілінің түсіндірме сөздігі" (1959, 1961 жж.) материалдары бойынша ЭЕМ арқылы қазақ тілі буындарының жиілік сөздігін жасады [9]. Ал Д.А.Байтанаева қазақ мәтіндеріндегі фонемалардың ғана емес сөздердегі буындардың да статистика-ақпараттық қасиеттерін тәжірибе жүзінде қарастырды [6].

Статистикалық әдіс түрлі тілдердегі грамматикалық тұлғаларды салыстыра зерттеуде де елеулі нәтижелерге ие болды. Мәселен, Т.М.Гарипов башқұрт, қазақ, татар, чуваш түркі тілдерін лексика-семантикалық тұрғыда талдап, олардың сандық және сапалық ерекшеліктерін анықтауды мақсат етті [27]. Сол сияқты туыстас емес тілдер қазақ және орыс, қазақ және ағылшын тілдерін салыстырғандағы грамматикалық формалар да статистикалық зерттеу аясынан тыс қалмады [4, 36, 51].

1987 жылы ҚазКСР "Наука" баспасынан шыққан (осы жолдар авторының) "Квантитативная структура казахского текста" және 2002 ж. жарық көрген "Основные принципы формализации содержания казахского текста" атты монографияларында қазақ тілінің заманға сай өзекті зерттеу мәселелері сөз етілді. Бірінші аталған ғылыми еңбекте қазақ тіліндегі мәтіндерді автоматты түрде электронды-есептеу машиналары (ЭЕМ) арқылы статистикалық әдіспен зерттеу мүмкіндігі мен сөз таптарының ықтималды-үлестірімді заңдылықтары зерттелсе, екінші монографияда – қазақ мәтінінің мәтін лингвистикасы тұрғысынан мазмұнын формальдау жолдары және мәтіннің туындау (порождение) проблемалық мәселелері қарастырылады [31, 39].

Сонымен, жоғарыда сөз болған ізденістер жеке автор шығармалары, функционалдық стильдер не толық әдеби тіл қамтылған мәтіндер бойынша статистикалық әдіс-тәсілдермен зерттеу тәжірибесінде әртүрлі жиілік сөздіктер жасаудың өзекті мәселе (актуалды) екендігін көрсетті. Себебі олардың негізінде аса маңызды және күрделі мәселелерді: тілді лексика-грамматикалық нормалау, түркі тілдерін өзара салыстырмалы-типологиялық зерттеу және олардың компьютерлік қорын жасауға зор мүмкіндік жасалады.

Статистикалық лингвистика саласына қатысты зерттеулер жайлы жоғарыда келтірілген қысқаша шолудан ең алдымен байқайтынымыз, мұндағы зерттеу аясының

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ
ҰЛТТЫҚ АКАДЕМИЯЛЫҚ КІТАПХАНАСЫ
№ 0018572

(нысанының) кеңдігі және бұл әдістің әмбебаптық сипаты. Солай бола тұра, біз қарастырған зерттеу жұмыстарының нәтижелері мен сапасы, деңгейі бірдей дәрежеде емес. Егер қазақ, қырғыз, әзірбайжан және өзбек тілдеріндегі ізденістер диссертациялық деңгейдегі іргелі зерттеулер қатарына жатса, ал кейбір басқа тілдерге қатысты статистикалық зерттеулер жаңадан ғана бастама алуда және олар жайлы тезис түріндегі қысқаша хабарламалар бойынша аз ғана пікір айтуға болады. Сол сияқты мұнда тіл білімінің барлық салалары да біркелкі қамтылмаған. Мәселен, қай түркі тілін алсаңыз да, олардың құрылымын статистикалық лингвистика тұрғысынан зерттеуде синтаксис пен семантика салалары тиісті дәрежеде қамтылмаған деуге болады. Әрине, аталған әдіс түркі тілдері үшін "жаңа" болғанымен, кейбір түркі тілдері үшін бұл әдіс тұрақтанып, түркітанушы ғалымдардың зерттеулерінен елеулі орын алған. Басқа түркі тілдерін айтпағанда, қазақ тілінің көптеген кезек күттірмейтін мәселелері, мәселен лексикография саласы бойынша компьютерлік қор (база) жасау, қазақ тілінен басқа тілге, не керісінше компьютерлік аударма жасау, ақпарат мазмұнынан қысқаша реферат, аннотация алу және қолданбалы лингвистика саласының басқа да мәселелері өз шешімін күтуде.

1.2. Жиілік сөздіктердің түрлері және олардың қолданбалы лингвистикадағы маңызы

Мәтін бірліктері және олардың сипаттамалары. Жазба мәтін – қағаз бетіне түскен жазба сөзі немесе соның бір үзіндісі, бөлігі. Мәтіндегі ой-пікірлер тек сөйлем түрінде ғана айтылғанда түсінікті бола алады. Ал сөйлем ойды айтып немесе жазып жеткізудің негізгі амалы. Сондықтан да мәтінді синтаксистік бірлік деп ұйғаруға болады.

Мәтіннің негізгі қызметі – жазбаша (не сөйлеу) түрінде жай немесе күрделі мазмұнды ой-пікірлерді білдіру. Бұл жағдайлардың бәрі де сөйлемдер тобы арқылы жүзеге асады. Олай болса, сөйлем – белгілі қоғам мүшелерінің өзара пікір алмасуын, қарым-қатынас жасауын қамтамасыз ететін мәтін бірлігінің біртұтас түрі.

Лингвистикалық атау (термин) сөздердің мағына-мәндерін бірыңғай ашып алу – тілді зерттеу жұмысының ғылыми және сапалы болуының куәсі. Осындай принципке негізделініп шығарылған ғылыми нәтижелер мен топшылаулар, тілдік заңдылықтар тиянақты да тұжырымды болады деп саналады. Алайда мәтін бойында кездесетін бірліктердің бәрін бірдей қарастыру қиындық туғызады. Сондықтан көптеген зерттеу жұмыстарында қолданылатын лингвистикалық атаулардың – сөз, сөзтұлға (сөзформа), сөзқолданыс, морфема, сөз және оның морфемалық, тұрпаттық құрылысы сияқты түрлерінің айырым белгі-қасиетін алдын ала білген жөн. Бұл атаулар мәтіннің (жазба не сөйлеу) не сөздіктердің бірлігі тұрғысынан қарастырылады.

Енді жиілік сөздіктер және олардың түрлері мен сипаттамаларына қысқаша тоқталайық.

"Сөз" деген атауды ақиқат шындықтағы зат пен құбылыстың атауы, сол атаудың дыбысталуы мен мағына-ұғымының бірлігі және семантикалық жағынан ақиқат шындықтың элементі деп түсінген жөн. "Сөз" атауы жөнінде айтыс туғызатын жаңа мәселелер бола тұрса да, соңғы кезде қазақ тіл білімі жаңа арнаға түсіп, дұрыс бағыт алып келеді. Біздің топшылауымызша, сөзтұлға дегеніміз – дербес

немесе көмекші мағыналы сөздің түбір тұлғасы мен оған жалғанған тұлға тудырушы морфемалардың сан алуан көріністері. Яғни түбір сөздің (негізгі тұлғаның) өзі де, оған сөз түрлендіруші грамматикалық форманттардың үстелуінен пайда болған тұлғалар да сол сөздің әртүрлі түрпаттары болып саналады.

Сөздің негізгі және туынды түрпатының, цифрлар мен әртүрлі символдардың және т.б. белгілердің мәтінде қайталанбай да, қайталанып та қолданылуы ***сөзқолданыс*** деп аталады.

"Сөз", "сөзтұлға", "негізгі тұлға" және ***"сөзқолданыс"*** деген атаулардың (терминдердің) лексикалық мағына жағынан өзара ерекшеліктері болғанымен, олар бір-бірімен тығыз байланыста болатын әртүрлі тілдік бірліктер.

Сонымен, сөз бен оның тұлғаларының және сөзқолданыстың мәтін бойындағы шекарасы – санауға жататын бірлік – екі ашық жер (пробел) арасындағы мәтіннің бөлігі (тыныс белгілерін есепке алмағанда). Бірақ мәтін бойында тек қана сөз және оның тұлғалары ғана емес, сонымен бірге цифрлар (сандық белгі) мен неше түрлі басқа да белгілер кездесуі мүмкін. Мәтіннің бұл элементтері де шартты түрде сөзқолданыс деп аталады да, олардың жиіліктерінің сөздікте көрініс табуы зерттеу мақсатына қатысты болады.

Тілдің стильдік салаларының лексикалық құрылымын статистикалық тәсілмен жүйелі түрде зерттеу мақсатымен жиілік сөздіктердің бірнеше түрін түзу (жасау) қажет болады.

Жиілік сөздіктердің көп тараған түрлері мынадай:

1) әліпби-жиілік сөздік;
2) жиілік сөздік;
3) кері әліпби-жиілік сөздік;
4) сөзнұсқағыш әліпби-жиілік сөздік (сирек құрастырылатын жиілік сөздіктің түрі).

Осы аталған жиілік сөздіктер үзінділері: *1.1, 1.2, 1.3, 1.4, 1.5-кестелерде* көрініс тапты.

Ескерту. Сөздікке М.О.Әуезов шығармаларының 14-томындағы 30–31-беттердегі мәтін үзіндісі негіз болды және ондағы жол саны кітап жолдарымен сәйкес келмейді. Аталған мәтін үзіндісі тақырыпша соңында берілді. Енді жиілік сөздіктерге қысқаша сипаттама берейік.

1.3. Әліпби-жиілік сөздік

Жиілік сөздіктің бұл түрінде зерттеу мақсатына сай таңдалып алынған мәтіндердегі сөздер (сөзтұлғалар) түгелімен қамтылуы қажет және олар бір-бірімен салыстырылып, әрқайсысының қолдану саны анықталады. Осындай сандық көрсеткіш сөздің (сөзтұлғаның) мәтін ішіндегі ***қолдану жиілігі*** деп немесе сөздің ***"абсолютті жиілігі"*** деп аталады.

Әліпби-жиілік сөздікте көрініс табатын ең бірінші (қатарда) – сөздік бірлігінің реттік сандары, екінші – мәтіннен бөлініп алынған әртүрлі сөздің (сөзтұлғаның) қатаң әліпби тәртібімен орналасқан тұлғасы және үшінші – сөздің абсолютті жиілігі.

Жиілік сөздіктердің толық түрінде осы аталғандардан басқа: "жиынтық абсолютті жиілік", "қатынастық жиілік", "жиынтық қатынастық жиілік" және сонымен бірге бір топ сөздердің мәтінді қамтуының пайыздық шамалары беріледі. Жиілік сөздіктер

жасау (түзу) арнайы компьютерлік бағдарламалар көмегімен іске асады. Сөздіктерді жасау барысының бастамасы сөздік бірлігінің әліпби тәртібіне келтірілмеген жиілікті сөзтізбе түрінде, яғни *1.1-кестеде* көрсетілген пішінде алынатынын ескеру қажет.

Осыдан кейін арнайы компьютерлік сұрыптау бағдарламасының негізінде басқа да сөздіктер түрлері түзіледі. Мәселен, әліпбилік сұрыптау арқылы *1.2-кестеде* көрсетілгендей "Әліпби-жиілік сөздік" алынады.

Әліпби-жиілік сөздік зерттеушіге қажет деген сөздің (оның туынды тұлғаларының да) қолдану дәрежесін, яғни оның жиілігін табуға, зерттеуге алынған мәтіндер лексикасының әліпбилік құрамының статистикасын анықтау үшін аса маңызды. Сондықтан, оны тілдің әліпбилік тұрғыдағы статистикалық үлгісі (моделі) деп те атауға болады.

Әліпби-жиілік сөздік, сөздіктің нормативті түріне жатпағандықтан, зерттеуші мәтіндегі немесе оның сөзтізбесіндегі сөздердің (сөзтұлғалардың) әдеби нормаға не орфографиялық жазылу нормасына жату-жатпауына жауап таба алмауы мүмкін. Оның себебі, сөздікте көрініс тапқан сөздер (не сөзтұлғалар), әдетте, қалам иесінің (автордың) не сол шығарманы жарыққа шығарушы жауапты адамның (редактордың) тіл игеру дәрежесіне қатысты да болады.

1.1-кесте (үзінді)

Әліпби тәртібіне келтірілмеген "жиілікті сөзтізбе"

СӨЗ (СӨЗТҰЛҒА)	Абсолютті жиілік	СӨЗ (СӨЗТҰЛҒА)	Абсолютті жиілік
1	2	1	2
Тіпті	1	ғана	4
түк	1	толас	1
білмейтін	1	тауып	1
бірақ	8	кішкене	2
оның	5	аласа	1
бұған	2	бөлмесінде	1
жұбаныштығы	1	болымсыз	3
аз	2	әңгімесін	1
болды	6	тыңдап	1
қашарын	1	варенье	1
білмеді	1	жеп	1
қыстақтан	1	қана	1
шықпай	1	отыратын	1
бір	3	Анна	3
жыл	1	Павловнаның	1
тұрғанда	1	күтуші	1
сол	6	қыздарының	1
мезгіл	1	ішінде	1
оған	2	көркем	1
он	1	қыз	5
жылға	1	өңі	1
бергісіз	1	нәзік	1
тек	2	көзінде	1
шешесінің	1	момындық	1
қасында	1	бар	4

1.2-кесте (үзінді)

Әліпби-жиілік сөздік

Рет саны (Р/с)	СӨЗ (СӨЗТҰЛҒА)	Абсолютті жиілік	Рет саны (Р/с)	СӨЗ (СӨЗТҰЛҒА)	Абсолютті жиілік
1	2	3	1	2	3
1	Андреевич	3	25	ақырын	1
2	Андреевичке	1	26	ақырып	1
3	Андреевичтің	1	27	ал	2
4	Анна	3	28	аласа	1
5	ағылшын	1	29	алғашқы	1
6	адам	1	30	алдампаз	1
7	адамдар	2	31	алмады	1
8	Дидерот	1	32	алмай	1
9	Дмитрий	2	33	алмайды	1
10	аз	2	34	алуандас	1
11	Иван	10	35	алуға	1
12	айғай	2	36	алып	3
13	айдап	1	37	алыс	1
14	айтқандай	1	38	алысы	1
15	айтты	2	39	Маланья	3
16	айтып	2	40	Маланьяға	1
17	айыбы	1	41	Маланьяны	2
18	айыбын	1	42	Марфа	1
19	айыптады	1	43	ант	1
20	айыпты	1	44	ап	1
21	ақ	1	45	Павловна	1
22	ақша	1	46	Павловнаның	1
23	ақылды	1	47	апара	1
24	ақыр аяғы	1	48	апарып	1

Тіл мәдениетін зерттеуші ғалымдар мұндай әліпби-жиілік сөздіктен бағалы мәліметтер таба алады. Сол сияқты, әр стильдік мәтіндердің өзіндік тіл ерекшеліктерін анықтауға және жалпы тіл біліміне қатысты көптеген теориялық мәселелердің басын ашуға мүмкіндік туады деуге болады.

Сөздік бірлігі ***"сөзтұлға"*** болып есептелетін әліпби-жиілік сөздік бойынша, бірлігі – ***"сөз"*** болатын екінші әліпби-жиілік сөздікке көшу қиынға соқтырмайды, себебі "сөзтұлға" сөздігінде бірдей лексемалар құрайтын сөзтұлғалар бірімен-бірі жалғаса орналасқандықтан олардың жалғаулық қосымшаларын оңай "қиып" тастау мүмкіндігі туындайды. Әліпби-жиілік сөздік үзіндісі *1.2-кестеде* көрсетілді.

Сонымен, әліпби-жиілік сөздіктердің бірлігі – сөз, сөзтұлға, түбірсөз, жай не фразеологиялық тіркес, морфема және т.б. бола алады және оларды жалпы атпен – "әліпби-жиілік тізім" немесе "жиілік тізім" деп атауға болады.

1.4. Жиілік сөздік

Сөздіктердің "жиілік сөздік" түрі әліпби-жиілік сөздіктен басқаша, дәлірек айтқанда, әр сөздің (сөзтұлғаның) қолдану жиілігінің дәрежесіне қарай орналасады: ең бірінші ретте орналасатын мәтіндегі ең жиі қолданылған сөз (не сөзтұлға), екінші, үшінші (тағы тағылар) ретте орналасатын сөздер – кездесу жиіліктері бірте-бірте кеміп отыратын сөздер (сөзтұлғалар). Егер бірнеше сөздің (не сөзтұлғаның) жиіліктері тұрақтылық сипатта болса, яғни біріне-бірі тең болып келсе, онда ол сөздер (сөзтұлғалар) жиілік сөздікте әліпби тәртібімен реттеледі. Осындай жиілік сөздіктің толық түрдегі үзінді тұрпаты *1.3-кестеде* көрініс тапты.

Зерттеуші (мұғалім, оқушы) әліпби-жиілік сөздік бойынша өзін "қызықтыратын" сөздің қандай жиілікте қолданылатынын немесе керісінше, алдын ала "қызықтыратын" жиілік шамасының қандай сөздерге тән екендігін анықтай алады.

Жиілік сөздік (әліпби-жиілік сөздіктер тәрізді) әртүрлі тілдік бірліктер (сөз, сөзтұлға, сөзтіркес т.б.) бойынша жасалуы мүмкін. Біздің баяндауымызда қарастыратын бірліктеріміз негізінен сөзге не сөзтұлғаға сәйкес келетіндіктен, "сөздің жиілік сөздігі" не "сөзтұлғаның жиілік сөздігі" деген сөз тіркестер жиі-жиі кездесуі мүмкін. Ал бұл сөздіктердің де жалпы аты – "жиілікті сөзтізбе".

Ана тіліміздің немесе басқа да туыстас (не туыстас емес) тілдердің әр стильдік мәтіндеріне қатысты жасалған жиілік сөздіктер сол тілдің лексикалық құрылымын терең зерттеуде, лексиканың жалпы функционалдық стильге ортақ бөлігін ажыратуда септігі мол.

Сондай-ақ, жалпы халықтық лексиканың, кірме сөздер мен терминдердің, неологизмдердің және басқа лексикалық топтардың тілдің жалпы лексика жүйесінен алатын орны жөнінде де жиілік сөздік бойынша көптеген сандық және сапалық мәліметтер алуға болады. Сөздердің саны мен тұлғалары арқылы тілдің лексикалық байлығы мен стильдік ерекшеліктері анықталады. Мұндай сөздіктерден зерттеуші жаңа сөзжасам жүйесіне қатысты да бағалы деректер ала алады. Сонымен бірге, жиілік сөздік мәліметтері қазақ тілін туыстас түркі тілдері және туыс емес тілдермен салыстыру – типологиялық зерттеу үшін де аса қажетті екенін атап өткен жөн [12].

Жиілік сөздіктер күнделікті қолданбалы сипаттағы міндеттерді шешуде де маңызды. Мәселен, әртүрлі типтегі минимум сөздіктерін жасауда және шет тілдерін (екінші тілді) үйрену мен оқытуға қажетті статистикалық әдістерді қолдануда да қажетті. Сонымен бірге жиілік сөздіктердің тілдік және сандық мәліметтері оқулықтар мен оқу құралдарын құрастыру ісінде тілдік материалдарды тиімді түрде орналастыруда және әрқилы лексикалық топтар мен грамматикалық категорияларды орынды түрде оқыту процесінде де таптырмас тілдік құрал рөлін атқара алады.

Жиілік сөздіктер материалдарының тіл білімін зерттеудегі қажеттілігі мен бағалылығы жөнінде Л.Н.Засорина мынадай пікір айтқан болатын: "Материалы частотных словарей исключительно ценны и для собственно лингвистических исследований. Они оказывают влияние на судьбы традиционной лексикографии. Ими пользуются в решении основной проблемы общей лексикологии – выделение словарного фонда активного и периферического словаря; они полезны и при изучении вопросов стилистики, семантики и литературной нормы языка" [41,3-4].

1.3-кесте

Жиілік сөздік (үзінді)

Рет саны	СӨЗ (СӨЗТҰЛҒА)	Абсолютті жиілік	Жиынтық абсолютті жиілік	Қатынасты жиілік	Жиынтық қатынасты жиілік	Бір топ сөздің мәтінді қамту пайызы
1	Иван	10	10	0.016077	0,016077	1,61
2	еді	9	19	0,014469	0,030547	3,05
3	бірақ	8	27	0,012862	0,043408	4,34
4	да	8	35	0,012862	0,056270	5,63
5	болды	6	41	0,009646	0,065916	6,59
6	сол	6	47	0,009646	0,075563	7,56
7	Петр	5	52	0,008038	0,083601	8,36
8	Петрович	5	57	0,008038	0,091640	9,16
9	әкесі	5	62	0,008038	0,099678	10,00
10	енді	5	67	0,008038	0,107717	10,77
11	қыз	5	72	0,008038	0,115756	11,57
12	ол	5	77	0,008038	0,123794	12,38
13	оның	5	82	0,008038	0,131833	13,18
14	Петровичке	4	86	0,006431	0,138264	13,83
15	баласына	4	90	0,006431	0,144695	14,47
16	бар	4	94	0,006431	0,151125	15,11
17	ғана	4	98	0,006431	0,157556	15,75
18	деп	4	102	0,006431	0,163987	16,40
19	өзінің	4	106	0,006431	0,170418	17,04
20	Андреевич	3	109	0,004823	0,175241	17,52
21	Анна	3	112	0,004823	0,180064	18,01
22	алып	3	115	0,004823	0,184887	18,49
23	Маланья	3	118	0,004823	0,189711	18,97
24	болатын	3	121	0,004823	0,194534	19,45
25	болымсыз	3	124	0,004823	0,199357	19,93
26	бұл	3	127	0,004823	0,204180	20,42
27	бір	3	130	0,004823	0,209003	20,90
28	деген	3	133	0,004823	0,213826	21,38
29	етті	3	136	0,004823	0,218650	21,86
30	жіберген	3	139	0,004823	0,223473	22,35
31	жігіт	3	142	0,004823	0,228296	22,83
32	мұның	3	145	0,004823	0,233119	23,31
33	адамдар	2	147	0,003215	0,236334	23,63
34	Дмитрий	2	149	0,003215	0,239550	23,95
35	аз	2	151	0,003215	0,242765	24,28
36	айғай	2	153	0,003215	0,245981	24,60
37	айтты	2	155	0,003215	0,249196	24,92
----	----------	-------	------	---------	---------	-------
78	шығып	2	237	0,003215	0,381029	38,10
79	шықты	2	239	0,003215	0,384244	38,42
80	ішінен	2	241	0,003215	0,387460	38,75
81	Андреевичке	1	242	0,001608	0,389068	38,91
----	----------	------	------	---------	---------	-------
460	ішік	1	621	0,001608	0,998392	99,84
461	ішінде	1	622	0,001608	1,000000	100,00

Жиілік сөздік өзге сөздіктердің көлеміне қатысты және басқа да лексикографиялық мәселелерді шешуде объективті негіз болатындықтан, екі немесе одан да көп тілдік аударма сөздік пен түсіндірме сөздік жасаушылар бұл сөздіктен қажетті деректер таба алады.

Жиілік сөздіктер телефон мен телеграф байланыстарын жетілдіру үшін де, стенография мен полиграфия мұқтаждықтары үшін де пайдаланылады.

Сайып келгенде, жалпы түркі тілдерінің, соның ішінде қазақ тілінің жиілік сөздіктерін жасау қажет деп саналады, өйткені олар көптеген лингвистикалық мәселелерге математикалық әдістерді қолданып зерттеуге мүмкіндік туғызады.

Қазақ тілінің сөз байлығын, яғни қазіргі қазақ әдеби тілінің жалпы лексикасын түгел қамтитын жиілік сөздік әлі жасала қойған жоқ. Мұндай сөздіктің құрастырылуы мен жарық көруі – жуық болашақтың ісі.

1.5. Кері әліпби-жиілік сөздік

Сөздіктердің басым көпшілігінде сөздер алдыңғы дыбыстар бойынша әліпби тәртібімен орналаса, кері әліпби сөздікте, керісінше, сөз бен сөзтұлғалар соңғы әріп-дыбысынан басталып әліпби тәртібіне келтіріледі және сөздердің соңғы жағынан тегістеліп беріледі. Сөздіктің бұл түрі "Кері әліпби-сөздік" немесе тек ***"Кері сөздік"*** деп аталып жүр. Егер белгілі бір мәтін бойынша түзілген мұндай сөздікте әр сөздің мәтіндегі қолдану жиіліктері де ескерілетін болса, ондай сөздікті "Кері әліпби-жиілік сөздік" деп атайды. Қазақ тілін зерттеу жұмыстарында мұндай сөздіктер де бірте-бірте қолданыс тауып келеді.

"Кері сөздіктің" алғашқы үлгісі осыдан бір ғасырдай бұрын (1873 ж.) баспадан шыға бастағаны белгілі. Бірақ біздің заманымызда бұл сөздіктерге қызығушылық тек XX ғасырдың 80-жылдарынан ғана басталды деуге болады. Мәселен, 1904 жылы латын тілі мен ескі иран тілінің кері сөздіктері жарияланды. 1944 жылы ескі грек тілінің кері сөздігі пайда болды. 1955 жылы ескі славян тілінің, ал 1957 жылы қазіргі румын тілінің кері сөздігі шықты. Чехославакияда 1935–1957 жылдары чех тілінің 9 томды сөздігі бойынша жасалған кері сөздік кішігірім кітапшалар түрінде басылып шықты.

Кері әліпби-жиілік сөздіктер жасауға айрықша көңіл бөлінудің нәтижесінде, Кеңес одағы кеңістігінде де әртүрлі тілдер бойынша осындай сөздіктер жарық көре бастады. Мысалы, орыс тіліне қатысты "Обратный словарь русского языка" (М., 1974), А.А.Зализняктың "Грамматический словарь русского языка" (М., 1977) және т.б.

Кері сөздіктердің рөлі, әсіресе, агглютинативті тілдер үшін ерекше екенін еске сала отыра, олардың саны әлі де болса санаулы. Осындай сөздіктердің алғашқылары ретінде: Қ.Бектаевтың "Қазақ тілінің кері алфавитті сөздігі" (Алматы, 1971), Р.Кунгуров, А.Тихонов "Обратный словарь узбекского языка" (Самарканд,1969, өзбек тілінде) және т.б. кері сөздіктерді атауға болар еді. Аталған сөздіктер мәтінге тікелей қатыссыз, яғни бұрын жарық көрген түсіндірме, орфографиялық және т.б. сөздіктер негізінде жасалғандықтан, оларда сөздердің қолдану жиілігі көрсетілмеген. Ал белгілі бір мәтінге байланысты жасалған кері сөздіктер грамматикалық форманттардың жұмсалуын зерттеуге бірден-бір қажет екені айқын.

Қазақ тілінің кері әліпби-жиілік сөздігінің ең алғаш жарық көруі М.О.Әуезов шығармаларының ЭЕМ арқылы зерттеле бастауымен тікелей байланысты. 1979 жылы "Ғылым" баспасынан шыққан "М.Әуезовтің "Абай жолы" романының жиілік сөздігінің" және 1995 жылы "М.О.Әуезовтің 20 томдық шығармалар

мәтіндерінің жиілік сөздіктерінің" жарық көргенін айрықша атап айтуға болады [13, 14]. Осымен қатар кейбір ғылыми, публицистикалық әдебиеттердің мәтіндері бойынша да ірілі-кішілі кері әліпби-жиілік сөздіктері жасалып, оларға біршама лингвостатистикалық талдау жасалды [58].

Жиілік сөздіктің кері әліпби түрінде реестрлік сөзге (сөзтұлғаға) қосылған біркелкі жалғаулар мен жаңа сөз жасаушы немесе сөз түрлендіруші жұрнақтар бір жерге жинақталып берілуі тілді зерттеушілер үшін аса құнды тілдік мәліметтер деп саналады. Кері әліпби-жиілік сөздік морфологиялық категорияларды зерттеу кезінде де және сөз тудырушы, сөз түрлендіруші морфемалардың құнарлы не құнарсыз екендігін анықтау үшін де, терминдер мен басқа тілдерден енген сөздерге жалғанған қосымшалардың құрылымын талдауда да аса бағалы. Осы айтылғандармен қатар, тіл зерттеушіге қажетті деген тілдік мәліметтерді әрі толық, әрі тез түрде тауып алу үшін кері әліпби-жиілік сөздік тілдік қазына деуге болады.

1.4-кесте (үзінді)

Кері әліпби-жиілік сөздік

Рет саны	Сөз (Сөзтұлға)	Абсолютті жиілік	Рет саны	Сөз (Сөзтұлға)	Абсолютті жиілік
1	2	3	1	2	3
1	қалаға	1	71	келуге	1
2	сахнаға	1	72	жіберуге	1
3	қораға	1	73	де	2
4	бақшаға	1	74	лезде	1
5	жолға	1	75	мүлде	1
6	жылға	1	76	білгенде	1
7	бұларға	1	-----	----------	------
8	тарихтарға	1	445	бірі	1
9	зорға	1	446	есі	1
10	дырдуға	1	447	әкесі	5
11	жазуға	1	448	келесі	1
12	алуға	1	449	шешесі	1
13	ұялтуға	1	450	кісі	1
14	Маланьяға	1	451	көршісі	1
15	да	8	452	тіпті	1
16	бойда	2	453	етті	3
17	тұрғанда	1	454	жетті	2
18	мұнда	1	455	кетті	1
19	қалпында	1	456	Петровичті	1
20	басында	2	457	үйленбекші	1
----	----------	-----	458	түземекші	1
68	үйге	2	459	елші	1
69	еденге	1	460	күтуші	1
70	бірге	1	461	Маланья	3

Кері әліпби-жиілік сөздіктердің лингвистикалық құндылығын И.Штиндлова өзінің мынадай ой-пікірімен білдіреді: "...Обратные словари могут быть использованы не только в качестве материала для освещения вопросов словообразования, не только как словари рифм или как пособия для восстановления испорченного текста, но они ценны, в первую очередь, при сравнительном изучении

родственных языков. И были бы интересны и для другого типа, например для аналитического языка, или для агглютинативных языков" [97].

Кері әліпби-жиілік сөздіктің әрбір жұрнақ пен жалғаудың (олардың вариантарымен) және әрбір сөз табының белгілі бір мәтіндердегі қолданылу жиіліктерін анықтауда мүмкіндігі мол екендігін тағы да атай отыра, осы аталған сөздіктің үзінді көрінісін *1.4-кесте* арқылы беруді жөн көрдік.

1.6. Сөзнұсқағыш әліпби-жиілік сөздік

Тіл білімінің лексикография саласындағы сөздіктер түзу мәселесі, әсіресе, түсіндірме сөздікті құрастыру барысына тән – ең алдымен неше түрлі сөздердің мағынасын ашатын картотекалық қор құру (жинақтау). Картотекалық қордағы әрбір карточкіде тілші-лексикографқа ең бірінші қажетті – "тірек сөз" (лексема) және, екінші ретте, сол сөздің мағынасына қатынасты мәнмәтін (контекст) келтіріледі. Осылармен бірге ол мәнмәтіннің қандай шығармадан алынғаны және оның шығу деректері: қалам иесі (автор), шығарма аты, баспа аты, шыққан жылы, көлемі, мәнмәтіннің кездескен бет саны да бірге берілуі керек.

Мәселен, А.Байтұрсынұлы атындағы Тіл білімі институтында көп жылдардан бері жинақталған осындай картотекалық қордың жалпы көлемі 5 млн карточкі-цитата (мәнмәтін) құрайды. Осы қордың негізінде 2 томдық, 10 томдық "Қазақ тілінің түсіндірме сөздігі" [59], "Абай тілі сөздігі" [1] және басқа да сөздіктер жарық көрді. Қазіргі кезде Институттың лексикограф-мамандары осы және басқа да тың картотекалық қордың негізінде – 15 томдық "Қазақ тілінің түсіндірме сөздігін" құрастыру үстінде.

Мұндай көлемді картотекалық қорды толықтыру, түзету немесе ұзақ мезгілге сақтау мен пайдалану жұмыстары тілші-лексикограф үшін оңайға түспейді. Сондықтан да, заманға сай, қазақ лексикография саласын автоматтандыру (немесе компьютерлендіру) кешенді проблемасын шешудің бастамасы ретінде А.Байтұрсынұлы атындағы Тіл білімі институтында "ТІЛ – ҚАЗЫНА" атты қазақ тілінің компьютерлік картотекалық базасы (қоры) іске қосылды. Оның негізгі мақсаты – қағаз бетіндегі 5 млн картотекалық қорды қайта сұрыптап, компьютер жадына, яғни "ТІЛ – ҚАЗЫНА" қорына енгізу және әрдайым ол қорды жаңа деректермен жаңартып отыру.

Осы аталған мәселені шешудің тағы бір жолы – "Сөзнұсқағыш әліпби-жиілік сөздігін" жасау. Бұл сөздіктің жоғарыда сөз болған "Әліпби-жиілік сөздіктен", "Жиілік сөздіктен" және "Кері әліпби-жиілік сөздіктен" негізгі айырмашылығы – баспадан шыққан шығарма мәтініндегі сөздердің кітап беттеріндегі орындарына нұсқау (мегзеу), яғни ол сөзтұлғаның (сөздің) қай бетте және сол беттің қай жолында кездесетініне сілтеме жасау. Сонда зерттеушіге қажет деген сөзге қатысты мәнмәтінді сол сілтеме арқылы кітап беттерінен оңай тауып алуына мүмкіндік туады. Қазақ тіл білімінде осындай әдіс тұңғыш рет А.Байтұрсынұлы атындағы Тіл білімі институтында М.О.Әуезовтің 20 томдық шығармалар жинағының жиілік сөздіктерін алу жағдайында қолданылды. Әрине, мұндай сөздіктің көлемінің үлкен болуы себепті, арнайы сөздік ретінде баспадан шығару оңайға түспесе керек. Ал ондай "Сөзнұсқағыш әліпби-жиілік сөздікті" тілші қауымының компьютерлік пішінде пайдалану мүмкіндігі мол. Аталған "Сөзнұсқағыш әліпби-жиілік сөздік" үзіндісімен *1.5-кесте* бойынша танысуға болады.

1.5-кесте (үзінді)

Сөзнұсқағыш әліпби-жиілік сөздік

Рет саны	Сөз (Сөзтұлға)	Абсолютті жиілік	Сөзнұсқағыш (n_i–m_j) (i – кітап беті, j – жол саны)
1	Андреевич	3	030-28; 031-11; 031-39
2	Андреевичке	1	030-20
3	Андреевичтің	1	030-41
4	Анна	3	030; 9; 030-26; 031; 4
5	ағылшын	1	031; 2
6	адам	1	031-22
7	адамдар	2	031-16; 031-36
8	Дидерот	1	030-33
9	Дмитрий	2	031-25; 031-26
10	аз	2	030; 3; 030-18
11	Иван	10	030-11; 030-15; 030-26; 030-30; 030-40; 031-1; 031-10; 031-15;
12	айғай	2	031-23; 031-28
13	айдап	1	030-24; 031; 4
14	айтқандай	1	031-14
15	айтты	2	031-22
16	айтып	2	030-35; 031-13
17	айыбы	1	030-40; 031-28
18	айыбын	1	030-37
19	айыптады	1	030-38
20	айыпты	1	030-30
21	ақ	1	030-36
22	ақша	1	031; 3
23	ақылды	1	031-21
24	ақыр аяғы	1	030-11
25	ақырын	1	030-19
26	ақырып	1	030-13
27	ал	2	031; 9
28	аласа	1	031-14; 031-25
29	алғашқы	1	030; 7
30	алдампаз	1	030-12
31	алмады	1	030-29
32	алмай	1	030-37
33	алмайды	1	030-42

Жоғарыда аты аталған тәжірибелік жиілік сөздіктеріне мәтіндік нысан болған М.О.Әуезовтің 20 томдық шығармалар жинағының 14-томынан екі беттік үзінді *(Аудармалар, 30–31-бб.):*

Французша тіпті түк білмейтін, бірақ оның бұған жұбаныштығы аз болды. Қашарын білмеді: қыстақтан шықпай бір жыл тұрғанда, сол мезгіл оған он жылға бергісіз болды. Тек шешесінің қасында ғана толас тауып, оның кішкене аласа бөлмесінде, болымсыз әңгімесін тыңдап, варенье жеп қана отыратын. Анна Павловнаның күтуші қыздарының ішінде бір көркем қыз болды, оның өңі нәзік, көзінде момындық бар, ақылды, Маланья дейтін қыз еді. Иван Петровичке сол қыз алғашқы көргеннен ұнады да, соған ғашық болды: қыздың сыпайы жүрісін, ұяң

жауабын, ақырын үнін, үнсіз жымиюын бұл түгел ұнатты: күн санап қыз ыстық көрінді. Иван Петровичке қыз да, барлық орыс қызындай, өзінің бар жанымен берілді. Шынымен қосылып еді. Қыстақтағы бай үйінде ешбір сыр, ұзақ уақытқа жасырын боп қала алмайды. Аз уақытта жас мырзамен Маланья жақындығын жұрттың бәрі білді; сол хабар ақыр-аяғы Петр Андреевичке де жетті. Өзге уақыт болса, әкесі мұны болымсыз іс деп елемес еді, бірақ баласына көптен зығыры қайнап жүрген әкесі енді Петербургтың білгішсіген сәнқой жігітін жерлеп, ұялтуға бекінді. Үлкен айғай, шаң-шұң шығып, Маланьяны апарып қараңғы үйге қамапты, Иван Петровичті әкесі шақыртыпты. Бұл дырдуға Анна Павловна да келген болатын. Ол күйеуін тоқтатпақ еді, бірақ Петр Андреевич енді сөз тыңдамайтын болды. Баласына қырғидай соқтығып, оны ұятсыз, алдампаз, құдайдан безген деп айыптады, Иван Петрович басында үндемей шыдап көріп еді, бірақ әкесі мұның намысына тиіп, жазалайтынын көрген соң шыдамады. Ішінен "құдайдан безген Дидерот тағы сахнаға шықты ма? Ендеше тоқтай тұрыңыз, мен оны іске жұмсап, сізді таңқалдырайын" деп ойлады. Сол әкесіне дау айтты; ұятсыз деген айыпты мойнына алмады; бірақ ол мүлде айыбы жоқ демеді; енді сол айыбын болымсыз сөздерді елемей, түземекші; қысқасы ол Маланьяға үйленбекші екенін білдірді; осы сөздерді айтып, Иван Петрович өзінің мұратына дәл жетті. Петр Андреевичтің жаңағы сөзге таңқалғандығы сонша, басында екі көзі бадырайып, үн шығара алмай қалды; бірақ артынан лезде есін жиып, үстіне тиін ішік киіп тұрған, жалаңаяқ аяғына башмақ киген қалпында жұдырығын түйіп, Иван Петровичке тап берді; жігіт бүгін әдейі де емес, шашын сыпайы тарап, үстіне ағылшын фрак киіп, шашақты етік, аса кербез ақ шалбар киген екен, Анна Павловна шошынғаннан айғай салып, бетін басты: жігіт зыта жөнелді, қашқан бойда бар үйден өтіп, қораға шығып, бақшаға қарай безіп, одан да өтіп жолға шықты: артына қарамай, зытып барып, әкесінің мұны қуған тысыры мен ауыр дем алысы басылғанша қашты. Әкесі ақырып: "Тоқта, бұзық, тоқта, теріс батамды берем", — деп еді, Иван Петрович тоқтамай кеткен бойда бір көшенің үйіне тығылды. Петр Андреевич қиналып, терлеп, демін зорға алып, үйге келе бере, баласына бата бермейтінін айтты; оның барлық құраған кітаптарын өртетпек болды; ал Маланья қызды қолма-қол алыс қыстаққа айдап жіберуге бұйрық етті. Иван Петровичке дос-жар адамдар табылып, оған жайдың бәрін мәлім етті. Ол ұялса да қатты ызаланып, әкесінен кек алуға ант етті. Сол түнде, крестьян арбасына мінгізіп апара жатқан Маланьяны жолшыбай тартып алып, қасына отырғызып, жақындағы қалаға қашып барды да, өздерінің некесін қиғызды. Бұған ақша берген көршісі үнемі ішкіліктен босамайтын, бірақ мейірімді адам болатын. Ол өзі айтқандай осы алуандас ғажайып тарихтарға жаны құмар кісі. Иван Петрович келесі күні әкесіне шанышпа тілмен, сыпайылау хат жазды, ал өзі шөберелес жақыны Дмитрий Пестовтың қыстағына кетті; Дмитрий Пестов бізге мәлім Марфа Тимофеевна атты қарындасымен бірге тұратын. Бұларға Иван Петрович өзінің бар жайын айтып, енді Петербургқа қызметке баратынын білдірген және мыналардан уақытша мұның әйелін қолдарына ұстауды сұраған. Әйелім деген жерде бұл жылап та жіберген, өзінің үлкен шаһардағы оқу төрбиесіне, философиясына қарамай, бишара ғана, мүләйім ғана орыстың бірі болып, туғандарының аяғына да жығылып бас иген, еденге маңдайын да тақ еткізген. Пестовтар қайырымды, мейірбан адамдар еді, мұның өтінішіне оңай ризалық білдірді; жігіт мұнда үш жұма тұрып, ішінен әкесінің жауабын күтіп еді, жауап келмеді, келуге мүмкін емес еді.

1.7. Мәтін мен оның жиілік сөздігі бірліктерінің арақатынасы

Тілдік бірліктердің деңгейі жоғарылаған сайын олардың формальды түрде жігін айыру қиындай түсетіні белгілі. Мұндай жағдай тіліміздегі лексика құрамының мағыналық аспектісінің көп жақтылық сипатына байланысты болады. Әсіресе, бұл сөздік қор бірліктерінің ішкі және сыртқы тілдік факторларының әсерінен бір-бірімен өзара тығыз және күрделі түрде әрекеттесуінің нәтижесінен болуы мүмкін. Сөздердің тілімізде қолданыс табуы кездейсоқтық оқиға деп есептесек, оның заңды не заңсыз құбылыс екендігі тек ықтималдықтар теориясы мен математикалық статистика әдістері арқылы ғана анықталады. Тек осы ғылымдар саласы ғана қайталанатын және біркелкілік сипаттағы оқиғалар аясымен шұғылдана алады. Сөздік қор мәселесін зерттеу барысында тілге әсер ететін қалам иесінің (автордың) әлемге деген көзқарасының барлық жақтары – әлеуметтік, тұрмыстық жағдай ерекшеліктері және т.б. бірдей ескерілуі тиіс. Себебі олар лексикалық қорға өз таңбасын қалдырумен бірге, тілді статистикалық жолмен зерттеудің қол жеткізер нысанына айналады.

Қазақ тілінің кейбір сандық заңдылықтар сапасының сырын ашуға көмектесетін деректерін, негізінен, зерттеуші-ғалымдар М.Әуезовтің "Абай жолы" романының мәтіні мен жиілік сөздігі негізінде қарастырды. Алынған деректер қазақ мәтінінің құрылымы жайлы дәстүрлі көзқарасқа да және жаңаша тұрғыдағы – статистикалық жолмен зерттеуге де көптеген мәліметтер беруде.

Зерттеу нәтижелері негізінде М.Әуезовтің "Абай жолы" романының мәтіні *465966 сөзқолданыстан* тұратыны анықталды. Ал ол романның әрбір томы (кітабы) бойынша есептелген сөзқолданыстар саны: 1-ші кітапта – *105788 сөзқолданыс*, 2-ші кітапта – *124398*, 3-ші кітапта – *112727*, 4-ші кітапта – *123053*. Бұл сандық шамалар романның әрбір кітаптарындағы сөзтұлғалардың қайталануларын қоса есептеу негізінде алынған мәтін көлемдері (ұзындығы).

Жоғарыда сөз болған жиілік сөздіктер түрін еске алсақ, олар – әліпби-жиілік сөздік, жиілік сөздік және кері әліпби-жиілік сөздік, сөзнұсқағыш әліпби-жиілік сөздіктер деп аталған болатын. "Абай жолы" романының осы аталған сөздіктер түрлері компьютер (ЭЕМ) көмегімен алынып, әрқайсысы бөлек-бөлек лингва-статистикалық зерттеу нысанына айналды.

Осы аталған сөздіктер ішінен арнайы қарастыратынымыз – "Жиілік сөздік" түрі. Оның анықтамасы бойынша, сөздердің орналасу тәртібі жиіліктерінің кему тәртібімен сәйкес келеді және бірдей жиілікті сөздер қатаң әліпби тәртібімен орналасады.

Егер мәтін бірлігі – **"сөзқолданыс"** термині арқылы аталса, сөздік бірлігі – ***"сөзтұлға"*** не ***"сөз"*** деп аталатынын еске түсіре кетейік.

Сонымен, компьютер көмегімен анықталған деректер бойынша "Абай жолы" романының әрбір сөздігіндегі *сөзтұлға* саны мынадай: 1-ші кітап сөздігінде – *22642 сөзтұлға*, 2-шіде – *26418*, 3-шіде – *26530*, 4-шіде – *27447 сөзтұлға*, ал толық роман бойынша (5-ші сөздікте) – *61824 сөзтұлға*.

Жазушы лексиконының байлығын сөз етудің ең қарапайым жолына қысқаша тоқталайық.

Егер мәтін көлемі – *"N"* деп белгіленсе, ал сөздік бойындағы сөзтұлға (не сөз) саны (сөздік ұзындығы) – *"L"* деп белгіленсе, онда $K_a=L:N$ – қатынас шамасы арқылы автордың сөздік қорының көп не аз екендігін қарапайым түрде, яғни

қатынас шамасының *"1"* санына жақын не алшақ болуына қарай пайымдауға болады [90, 42-43].

Мәселен, $K_a=L:N$ қатынасы *1*-ге тең болса, онда бөлшектің алымы бөліміне тең деген сөз, ал мұндай теңдік орындалу үшін мәтін тек қайталанбайтын сөздер тізбегінен ғана тұруы қажет. Бұл жағдай, әрине, көркем шығармалар мәтініне тән құбылыс емес. Егер де қатынастың алымы бөлімінен аз айырмашылықта болса қатынас шамасы да *"1"* санынан жуық болады, яғни автордың сөздік қоры бай, ал керісінше болса, яғни бөлшектің алымы бөлімінен көп кіші, онда шығарманың (қалам иесінің) сөздік қоры "кедей" деп пайымдалады.

Қазақ тілінің көркем әдебиет стилі ("Абай жолы" романы) және ағылшын, орыс, латын тілдерінің публицистика стильдерінің сөздіктері негізінде есептелген $K_a=L:N$ қатынас шамалары *1.6-кестедегідей* көрініс тапты.

Көптеген зерттеу деректері бойынша байқалатын жайт, $K_a=L:N$ қатынасының мәтін көлемінің құбылуына қарай өзгеріске ұшырауы және жуық көлемдегі мәтіндерде ондай қатынас мәндерінің де шамалас болуы. Сондықтан ескерту ретінде айтарымыз, мәселен, екі не үш жазушының шығармалар тілін зерттеу қажет болатын болса, олардан алынатын мәтіндер көлемі де бір-бірімен жуық шамада болуы шарт. Сол сияқты, олардан түзілетін сөздіктер бірліктері де бірдей дәрежеде болуы да шарт, яғни олар не сөз, не сөзтұлға түрінде болуы керек.

Жоғарыда сөз болған $K_a=L:N$ коэффициенті тілдің аналитикалық не синтетикалық сипатын анықтау үшін де қолданылады. Бұл коэффициент мәні неғұрлым "бірге" жуық болса, соншалықты ол аналитикалық тілге жақын да, ал егер ол "нольге" жуық шама болса, онда ол тілдің синтетикалық сипаты арта түседі деп пайымдалады. Түрлі тілдерге тән мұндай қасиетті *1.6-кестеден* байқауға болады. Кесте деректері бойынша аналитикалық-аморфты ағылшын тілінде K_a коэффициентінің мәні ең үлкен шама – 0,62 пайыз, агглютинативті қазақ тілінде ең аз шама – 0,28 пайыз. Ал флективті-синтетикалық орыс, латын тілдерінде K_a мәні ағылшын тілінен аз да, ал агглютинативті қазақ тілінен көп. Сонымен, бұл коэффициентті $(K_a=L:N)$ тілдердің аналитикалық (синтетикалық) сандық сипаттамасы ретінде де алуға болады деп қорытындылауға болады.

1.6-кесте

Тілдердің аналитикалық не синтетикалық сипатын K_a коэффициенті арқылы анықтау

Тіл	Стиль	Сөздер саны	Сөзтұлға саны	K_a
Ағылшын тілі	публиц.	5197	8300	0, 62
Орыс тілі	публиц.	5225	14117	0, 37
Латын тілі	публиц.	13319	45211	0, 30
Қазақ тілі: "Абай жолы"	көркем әдебиет	16983	61424	0, 28

Сөздік пен мәтін арақатынасын анықтаудың ең бір қолданбалылық маңызы зор түрі – жиілік сөздіктегі сөздің не сөздер тобының мәтін бойын қамту дәрежесін айқындау. Мәселен, шет тілін не ана тілін үйренуде, не үйретуде минимум сөздіктер жасау қажеттігі туатыны белгілі. Минимум сөздікке енетін сөздер сан жағынан аз болғанымен, мәтін ішінде көп қайталанып барып, оның өн бойын қамту мүмкіндігі

мол. Немесе тілімізде кейбір сөздер сирек қолданылғанымен, ондай сөздерді ескермеу, мәтін мазмұнының (сөйлем мағынасының) дұрыс ашылмауына әкеліп соғуы ықтимал. Міне, осындай және басқа да қолданбалы мәні жоғары мәтін мен сөздік арасындағы қатынастар туралы кейбір зерттеулерден мәліметтер келтіре отырып, сөздердің мәтінді қамту сипатына тоқталайық.

Ол үшін біз, өз баяндауымызда, "Абай жолы" романының жиілік сөздіктері мен мәтініне қатысты кейбір мәліметтеріне жүгінеміз.

М.Әуезовтің "Абай жолы" романында қанша сөз немесе сөзтұлға бар және олар қандай жиілікте қолданылған, мәтін мен жиілік сөздіктер бірліктерінің арақатынасы қандай деген мәселелердің басын ашу тек жиілік сөздіктер негізінде ғана іске асады. Жиілік сөздіктердің статистикалық сипаттамалары тілдің теориялық және қолданбалы мәнін неғұрлым тереңірек танып-білуге мүмкіндік беретіні анық. Жиілік сөздіктерді олардың бірлігіне (сөзтұлға және сөз) қарай екі жағдайда: сөзтұлға негізіндегі жиілік сөздік және реестрлік сөз (лексема) негізіндегі жиілік сөздік ретінде қарастыруға болады. Мәтін бойындағы сөзқолданыстардан компьютер арқылы жасалатын әліпби-жиілік сөздіктің бірлігі – сөзтұлға. Маман-лексикограф мұндай сөзтұлға сөздікті өз қалауынша бірлігі "сөз" болатын сөздікке (қосымшаларын "қию" арқылы), яғни реестрлік сөз түріне келтіруіне мүмкіндігі бар.

Жиілік сөздіктің құрылымы бойынша жиі қолданыстағы сөздер (сөзтұлғалар) сөздік бойының бас жағында орналасатындығын ескеріп, олардың мәтін мен сөздікті қамту пайызы жайлы мәліметтер *1.7-кестеде* берілді.

1.7-кесте

Бір топ сөзтұлғаның "Абай жолы" романы мәтіні мен сөздік бойын қамту пайызы

Бөліктер рет саны	Рангтық шекаралық	Мәтін бойын қамту пайызы	Сөздік бойын қамту пайызы
1	1	1	0, 002
2	1– 5	5	0, 008
3	1 – 16	10	0, 026
4	1 – 33	15	0, 053
5	1 – 62	20	0, 100
6	1 – 104	25	0, 168
7	1 – 168	30	0, 272
8	1 – 259	35	0, 419
9	1 – 391	40	0, 632
10	1 – 842	50	1, 362
11	1 – 1749	60	2, 829
12	1 – 3731	70	6, 035
13	1 – 8430	80	13, 635
14	1 – 61824	100	100

Көңіл бөлетін жайт "Абай жолы" романындағы ең жоғарғы жиілікті бірінші реттегі сөзтұлға мәтін көлемінің *1%,* ал сөздік бойының тек қана *0,002* пайызын қамтитыны. Сол сияқты, жиі қолданыстағы *5* сөзтұлға мәтіннің де *5* пайызын, ал олар сөздік бойының *0,008* пайызын ғана қамтитынына сөздік мәліметтері бойынша

көз жеткізуге болады. Сөздік бойының *1,4* пайызын құрайтын *842* сөзтұлға, мәтін ішінде қайталанып қолданудың негізінде мәтіннің *50* пайызын қамтиды екен, яғни бұл романның тең жартысы – *232983* сөзқолданыс. Осы тәріздес басқа мәліметтер де *1.7-кестеде* көрініс тапты.

Енді сөздік бірлігі сөзтұлғадан "сөзге" ауысқан жағдайдағы сөздердің түрлі шығармаларда қолданылу жайындағы статистикалық деректеріне тоқталайық. Мәселен, Абай тілінің жиілік сөздігі бойынша Абай шығармалары тіліндегі ең жиі қолданылған *75* сөз барлық сөзқолданыстың *40,8* пайызын, *150 сөз 50,8* пайызын, *1000* сөз *80* пайызын құрайды екен. "Абай тілі жиілік сөздігі" арқылы аталған қолданыстағы сөздердің лингвистикалық табиғатын да, яғни тілдік сыр-сипатын да ашуға болады.

Қолданылу жиілігі *500-ден* жоғары болып кездесетін "сөз" саны "Абай жолы" романда – *144.* Бұл жалпы сөз санының *0,84* пайызын құрайды, ал олардың қолданылу жиілігі жалпы сөз қолданыстың *48* пайызын құрайды. Бұл сандық сипаттама үндіеуропа тілдерінен айтарлықтай алшақ емес. Шынында да, көлемі шамалас кез келген түркі және үндіеуропа тілдеріндегі мәтіндерді алып, ең жиі қолданылатын бір мың сөздің мәтін қамтуының сандық мәнін салыстырсақ, бұл түркі тілдерінде *63* пайыз бен *86* пайыз аралығында, ал үндіеуропа тілдерінде *64* пайыз бен *89* пайыз аралығында екендігі анықталды [12]. "Абай жолы" романындағы ең жиі қолданылған *1000* сөз (сөзтұлға емес) мәтіннің *77* пайызын қамтиды. Сонымен, бұл сандық сипаттама тілдік универсалия (әмбебаптық) міндетін атқаратындығының айғағы.

Бірдей жиілікті сирек қолданатын сөздерді (сөзтұлғаларды) жеке қарастырсақ, ең бірінші байқалатын жайт – жиілігі кеміген сайын ондай сөздердің (сөзтұлғалардың) санының да өсуі. Мәселен, ең сирек қолданыстағы сөздердің (сөзтұлғалардың) кездесу жиілігінің кемуі (азаюы), олардың сөздіктің бірлігі ретіндегі сандық мөлшерінің ұлғаюына да әкеліп соғады екен. Сондықтан, *1* рет қолданыста болатын сөздердің саны *2* рет қолданыстан әрі көп, әрі мәтін мен сөздік бойын қамту жағынан да алдыда тұрады. Сол сияқты, жиіліктері *2* реттен болатын сөздер саны мен мәтінді қамту мүмкіндігі *3* реттегіге қарағанда көптік сипатта. Осы айтылғандар *1.8-кестеде* көрініс тапты.

Сирек қолданыстағы сөздердің статистикалық деректері жайлы "Абай жолы" романы мәтінімен бірге басқа да шығармаларды қарастырайық. Мәселен, газет тіліндегі тек бір реттен қолданылған *5720* сөз (омонимдер ажыратылды) сөздіктегі барлық реестрлік сөздердің *(12424) 46,04* пайызын, ал мәтіндегі барлық сөзқолданыс санының *3,81* пайызын құрайтыны анықталды. Ал Абай шығармалары тілінің бір реттен жұмсалған *2975* сөз барлық сөздіктегі реестрлік сөздердің *(6017) 49,4* пайызын, ал жалпы сөзқолданыс көлемінің *6,35* пайызын құрайтыны анықталды. Бұл фактінің өзі тек публицистика тілі мен көркем шығармалар тілінің ерекшеліктерін ғана емес, Абайдың көркем сөздің шебері және тілінің соншалықты бай екендігін тағы да дәлелдей түскендей.

"Абай жолы" романның әліпби-жиілік сөздігі реестрінде де және жиілік сөздік реестрінде де бірдей *16983* сөз орналасқан. Бұл романда осынша әртүрлі сөз бар (қолданылған) дегенді де білдіреді.

"Абай жолы" романында (*1.9-кесте*) тек *1* және *2* рет қана қолданылған *8698* сөз барлық *16983* сөздің *50* пайызын құраса [13], Пушкин шығармаларында осындай жиіліктегі *9301* сөз барлық *21197* сөздің *44* пайызын [100], "Абай тілі сөздігі" бойынша *3877* сөз барлық *6017* сөздің *64* пайызын [1], Мамин-Сибиряктің

"Приваловские миллионы" шығармасындағы *7312* сөз барлық *11283* сөздің *65* пайызын [101] құрайды. Мұндай сирек қолданылатын сөздердің сандық сипаттамасы жазушының тіл байлығы туралы мәлімет береді.

Жоғарыда аталған шығармалардың сөзқолданыс көлемі әртүрлі. Бірдей көлемдегі мәтіндерде сирек қолданылған сөздердің статистикалық сипаттамаларын салыстырып зерттеу арқылы құнды нәтижелер алуға болатыны сөзсіз. Ал ұлы жазушы М.О.Әуезовтің "Абай жолы" роман-эпопеясы бойынша жасалған жиілік сөздіктер деректері тек қазақ тілінің ғана емес, басқа түркі тілдерінің де ақын-жазушыларының сөздік қорын статистикалық лингвистика тұрғысынан зерттеуге бастама және түрткі болуы мүмкін.

1.8-кесте

"Абай жолы" романы мәтіні және жиілік сөздігі бойынша бірдей жиілікті сөзтұлғалардың қамту пайызы

Сөзтұлға жиілігі	Реестрдегі сөзтұлға саны	Сөздік бойын қамту пайызы	Жиынтық абсолютті жиілік	Мәтін бойын қамту пайызы
1	34860	56, 3	34860	7, 48
2	8739	11, 4	17478	3, 75
3	4261	6, 89	12783	2, 74
4	2540	4, 11	10160	2, 18
5	1810	2, 93	9050	1, 94
6	1184	1, 92	7404	1, 52
7	937	1, 52	6559	1, 41
8	785	1, 27	6280	1, 35
9	638	1, 03	5742	1, 23
10	499	0, 81	4990	1, 07

Енді М.О.Әуезовтің 20 томдық шығармалар жинағындағы 2,1 млн сөзқолданыстан тұратын мәтіні мен 29483 реестрлік сөзден тұратын (жалқы есімді қоспағанда) жиілік сөздігі арасындағы сандық қатынастарға қысқаша тоқталайық. Мұндай сандық қатынастар *1.10-кестеде* көрсетілді.

Жиілік сөздіктегі сөздер жиілігіне сәйкес топ-топқа бөлініп, олардың тізбедегі сөздердің қанша пайызын және қайталану қабілетін ескеріп, мәтіннің (жалқы есімсіз) қанша пайызын қамтитынын аталған кестеден (*1.10)* байқауға болады. Мысалы, жиілігі *10000*-нан астам *17* сөз тізбедегі (реестрдегі) барлық сөздердің *0,0006%* қамтығанымен, олар қайталана келе бүкіл мәтіндегі сөзқолданыстың *18,26%* құрайды. Сол сияқты, жиілігі *5001*-ден *10000*-ға дейінгі сөздердің саны 30-ға тең. Тізбедегі *29483* сөздің бұл *0,001%* пайызы, ал барлық мәтіннің (қайталануын ескергенде) не бары *0,122%*-ы екен. Жиілігіне сәйкес басқа топтардың қамту пайыздарын осылайша *1.10-кестеден* анықтауға болады.

Осы кестеден (не болмаса жиілік сөздіктің өзінен) барлық тілдерге (қазақ тілінде де) тән заңдылық байқалады. Ол – сөздердің жиілігі кеміген сайын реестердегі сөз санының ұлғаюы, бірақ мәтіндегі сөзқолданысты қамту пайызы барынша аз болуы. Бұл заңдылықты кестедегі жиілігі *10*-нан *1*-ге дейін кеміп отыратын сөздердің санынан анық байқаймыз. Мысалы, жиілігі *10*-ға тең сөздердің саны сөздікте *378 (0,013%)* болса, жиілігі *9*-ға тең сөздердікі – *465 (0,016%)*. Осылай, жиілігі кеміген

сайын олардың қатары арта түсетінін аңғарамыз. Мәселен, жиілігі *2*-ге тең сөз саны *3712 (0,126%)*, ал жиілігі *1*-ге тең болғанда – *10 988 (0,373%)* дейін өсетіні заңды құбылыс деп есептеуіміз қажет. Мұндай төменгі жиілікте қолданған сөздер жазушы тілінің мәйегі деуге болады.

1.9-кесте

"Абай жолы" романы жиілік сөздігі бойынша жиілікке сәйкес реестрлік сөз және сөзтұлға сандарының арақатынастары

Реестрлік бірліктің жиілігі	Жиілікке сәйкес сөз саны	Жиілікке сәйкес сөзтұлға саны	Реестрлік бірліктің жиілігі	Жиілікке сәйкес сөз саны	Жиілікке сәйкес сөзтұлға саны
1	2	3	1	2	3
2000-нан жоғары	27	11	11-20	1357	2525
1001-2000	33	23	10	219	499
501-1000	84	60	9	281	638
401-500	28	30	8	331	785
301-400	65	46	7	384	937
201-300	115	112	6	537	1184
101-200	332	318	5	680	1810
51-100	548	661	4	903	2540
41-50	260	350	3	1362	4261
31-40	353	480	2	2329	8739
21-30	591	955	1	6956	34460

Көбіне жазушының сөз байлығын осындай сирек кездесетін сөздердің санымен мөлшерлейді. Шынында да, М.Әуезовтің тың сөз қолдану шеберлігі оқырмандарға белгілі, сонымен қатар бір қолданған сөзін екінші рет қайталамай, баламасын іздеп синоним, дублет сөздерді көп қолданатыны да сөз байлығын көрсететін бірден-бір дерек. Әрине, басқа да белгілі сөз шеберлерінің жиілік сөздіктері жасалса, олардағы осы құбылыстарды айқындап, салыстыра зерттесе, әр автордың өзіне ғана тән тіл ерекшеліктерін нақты айтуға болар еді. Мұндай зерттеулер болашақтың ісі.

Енді *2,1* млн сөзқолданыстан тұратын М.Әуезовтің 20 томдық шығармалар жинағындағы сөз таптарының сөздік пен мәтін арасындағы салыстырмалы статистикалық деректеріне қысқаша тоқталамыз.

1.11-кестедегі сандық көрсеткіштен мәтінде ең көп қайталанып қолданылатын сөз таптары алдымен етістік, екінші орында – зат есім, ал ең аз кездесетіндер одағай мен еліктеу сөздер.

Мәтіндегі сөзқолданыстың *18%*-ын құрайтын жалқы есімдерді қоспағанда М.Әуезовтің 20 томдық шығармалар жинағында *29483* сөз қолданылғаны анықталды. Оның *5743*-і кірме сөздер. Мұндағы кірме сөздер – негізінен орыс тіліндегі орфограммасын сақтап енген сөздер және солардан өрбіген туынды түбірлер. Бұлардың қатарында қазіргі орфографиялық нормаға сәйкес келмейтін араб-парсы сөздері де бар.

Мысалы, *фарсы, фиғыл, фарасат, фақыр, фәле* т.б. Орыс сөздері де әртүрлі фонетикалық нұсқада ұшырайды. Мысалы, *прокурор - пыркорол, протокол – пүртөкөл, генерал – жанарал, приговор – пригоуор – пірғауар* т.б. Кейіпкерлердің тіліндегі айтылу ыңғайына орай жазылған әртүрлі фонетикалық нұсқадағы сөздерге де тізбеде (реестрде) орын берілді. Мысалы, *ажуа* (*55* рет қайталанып қолданылған), *әжуа – 22, әжуала – 7, ажым – 65, әжім – 14, ажымды – 11, әжімді – 4* т.б.

Бұл кестеден М.Әуезовтің сөз байлығының үштен екісі зат есім мен етістік сөздер екенін байқаймыз. Оның есесіне есімдік пен шылау сөздер бірігіп, жазушы тілінің жарым пайызын құрайды, алайда олардың мәтінде қайталану жиілігі өте жоғары, барлық сөзқолданыстың *18%*-ына жуықтайды.

Мәселен, зат есім тудыратын ең өнімді ***-шы, -ші*** жұрнағы *343* сөздің құрамында кездессе, етістік тудыратын – ***ла//-ле, -да//-де, -та//-те*** *1409* сөз жасап тұр.

Жазушы тілінде кездесетін сын есім сөздердің жартысынан көбі *(3054* сөз*)* ***-дай//-дей, -тай//-тей, -лы//-лі, -ды//-ді, -ты//-ті, -сыз//-сіз*** жұрнақтары арқылы жасалған.

1.10-кесте

М.О.Әуезовтің 20 томдық шығармалар жинағы мәтіні мен жиілік сөздігі арасындағы сандық қатынастар

Реестрлік сөз жиілігі (аралығы)	Жиілікке сәйкес сөз саны	Жиынтық абсолютті жиілік	Сөздікті қамту пайызы	Мәтінді қамту пайызы
1	2	3	1	2
10000-жоғары	17	311559	0, 0006	18,260
5001-10000	30	208325	0,001	0,122
1001-5000	217	443576	0,007	0,260
501-1000	244	171572	0,008	0,101
401-500	120	53964	0,004	0,032
301-400	172	60588	0,006	0,035
201-300	401	97836	0,014	0,057
101-200	839	118999	0,028	0,070
51-100	1091	78077	0,037	0,046
41-50	454	20516	0,015	0,012
31-40	712	25014	0,024	0,015
21-30	1056	26429	0,036	0,896
11-20	2370	33602	0,080	0,097
10	378	3780	0,013	0,002
9	465	4185	0,016	0,002
8	512	4096	0,017	0,002
7	659	4613	0,022	0,003
6	762	4572	0,026	0,003
5	992	4960	0,034	0,003
4	1364	5456	0,046	0,003
3	2028	6084	0,069	0,004
2	3712	7424	0,126	0,004
1	10988	10988	0,373	0,006
Қосындысы:	29483	1706195		

1.11-кесте

М.Әуезовтің 20 томдық шығармалар жинағындағы сөз таптарының сөздік пен мәтін арасындағы сандық қатынасы

Рет саны	Сөз табы	Сөз саны	Сөздік бойындағы үлесі (%)	Сөзқолданыс саны	Мәтін бойындағы үлесі (%)
1	Зат есім (зт)	8843	37, 25	520521	30, 50
2	Етістік (ет)	6853	28, 87	534836	31, 35
3	Сын есім (сн)	5885	24, 79	189224	11, 00
4	Есімдік (ес)	209	0, 88	173495	10, 17
5	Сан есім (са)	94	0, 40	28009	1, 64
6	Үстеу (үс)	1023	4, 31	74472	4, 36
7	Шылау (шл)	162	0, 68	129536	7, 60
8	Одағай (од)	327	1, 38	8408	0, 50
9	Еліктеуіш (ел)	344	1, 45	2029	0, 12
	Кірме сөздер (бн)	5743	19, 40	45265	2, 65
	Қосындысы:	29483	***Жалқы есім мен кірме сөздерсіз***	1,70619	***Жалқы есімсіз***

М.Әуезовтің шығармалар жинағында жалқы есімдер *386729* рет қайталанып қолданылған. Бұл сөздер сөздіктің реестр қатарында жоқ, бірақ әр том бойынша және түгел 20 том ішіндегі жалқы есім жайлы мәліметтер келтірілген. Жалқы есімдер өзара сегіз түрге бөлініп, мәтінде шартты белгілермен таңбаланады. 20 томдық шығармалар жинағы бірнеше жанрда, әртүрлі стильмен жазылғаны белгілі. Осыған байланысты жалқы есімнің түрлері барлық томдарда бірдей емес. Мысалы, пьеса мен драмалық шығармаларда адам есімдері көп кездессе, әңгіме, повесть, роман мәтіндерінде ру, тайпа және жан-жануар аттары көбірек кездеседі.

386729 рет қайталанған 20 томдық шығармадағы жалқы есімдер жайлы мәліметтер төмендегідей: адам есімдері – *173300* рет, жер-су аттары – *115396* рет, ру-тайпа аттары – *89705* рет, жан-жануар аттары – *3813* рет, баспа аттары – *3345* рет, мекеме аттары – *592* рет, қысқарған сөздер – *526* рет және атақ-даңқ аттары – *12* рет қайталанып қолданылған.

Бұл мәліметтерден бірден көзге түсетіні – мәтіннің өн бойында адам аттары мен есімдік сөздердің қайталану санының *(173300; 173495)* жуық болуы және атақ-даңқ, лауазым атауларының сирек қолданылып, тек мақалалар мен зерттеулерде ғана аздап кездесетіндігі.

Сонымен, М.Әуезовтің 20 томдық шығармалар жинағында қайталанып, әртүрлі сөз таптары мен атаулардың сан мөлшерінің өзіндік қыры-сыры, тілдік танымтүсінігі бар.

Қорыта келе айтатынымыз, М.Әуезовтің 20 томдық шығармалар жинағының үш түрлі жиілік сөздіктері жазушының тіл байлығы мен стильдік ерекшеліктерін тереңірек зерттеуге аса құнды-құнды материал болатындығы сөзсіз.

Жеке қалам шеберлерінің сөздігімен қатар, қазақ тілінің сөз байлығын, яғни қазіргі қазақ тілінің жалпы лексикасын түгел қамтитын жиілік сөздік құрастыру да күн тәртібінде тұрған мәселе – заман қажеттілігі. М.Әуезовтің 20 томдық

шығармалар жинағының жиілік сөздіктері болашақта жасалатын толық "Қазақ тілінің жиілік сөздігіне" қосылатын сүбелі үлес болатыны анық.

1.8. Жоғары жиілікті сөздердің лингвистикалық табиғаты

М.О.Әуезовтің "Абай жолы" романының барлық кітабында, шартты түрде алғанда, "аса жиі қолданылатын аймақ" деп аталатын сөзтұлғалардың жиынтығы лингвистикалық заңдылық тұрғысынан қарастырылды. Біз осы аталған романның жиілік сөздігіндегі жиі қолданыстағы 100 сөзтұлғаны бөліп алып, оларға статистика-лингвистикалық тұрғыда сипаттама беруді мақсат еттік.

Морфологиялық құрамына қарай жиі қолданыстағы 100 сөзтұлға әртүрлі топтарға бөлінеді. Мәселен, жалғаулық шылаулар ***да//де, мен//пен, және;*** сұраулық ***ма, ме,*** сондай-ақ *соң, сияқты, қарай, үшін, ал, ғана, бірақ, тағы, ғой* сияқты сөз байланыстырушылар көмекші сөздерге жатады. Олар тілде жеке бірлік ретінде қолданылмайды, тек синтаксистік байланыс құралдары ретінде қызмет етеді де, тілдегі басқа элементтермен бірге жиі қолданылатындардың қатарына кіреді. Жазушының бұларды жиі қолдануының себебін қазақ тіліндегі сөздердің қатынасқа осы көмекші сөздердің көмегімен түсетіндігімен түсіндіруге болады.

Бұл шығармада жиі қолданылған етістіктер бөліп қарастыруды қажет етеді. Етістік негізді сөзтұлғаларға *е* және *де* түбірлерінен өрбіген *еді, екен, етін,* сондай-ақ *деп, деді, деген* сияқты көмекші етістіктер жатады. Аталған екі етістік те, яғни олардан жасалған көсемше, есімше тұлғалары да статистикалық қызметіне қарай жоғарыда аталған көмекші сөздер сияқты әртүрлі синтаксистік құрылымдарды ұйымдастыруда, мынадай модельдегі етістік+етістік, есім+етістік және басыңқы мен бағыныңқы сөйлемдерді байланыстыруда көмекші етістік қызметін атқарады. Және де *"де"* етістігінен жасалған тұлғалар құрмалас сөйлем ішіндегі төл сөздің ажырамас бөлігі ретінде де қолданылады. *"Е"* және *"де"* етістіктерінің статистикалық жұмсалымдығы оларды жиі қолданылатын тұлғалар қатарына қосып отыр. Қазақ грамматикасының заңдылығына сүйенсек, етістіктің *"е"* түбірі бұйрық рай мағынасында берілуі керек еді, бірақ қазіргі қазақ тілінде ол өзінің бастапқы лексикалық мағынасын жоғалтып алған да, енді қосымша аналитикалық формантың құрамында ғана қолданыла алады. Бірақ етістіктің *"е"* түбірінің бастапқы мағынасын жоғалтуы оның белсенді қолданылуына шек қойып отырған жоқ. Оған оның романдағы аса жиі қолданылысы дәлел бола алады.

Аса жиі қолданылатын аймаққа кезінде, жарты ғасыр бұрын профессор Қ.Қ.Жұбанов алғашқы болып теориялық сипатын көрсетіп берген төрт көмекші етістік те кіреді. Әлі түрлері толық анықталып болмаған барлық күрделі етістіктердің ішінен өте жиі қолданылғыштығы мен сан қырлылығы жағынан *отыр, жүр, тұр* және *жатыр* деген төрт етістікті бөліп айтуға болады. Бұл төрт етістік бір топқа кіріп, басқалардан үш морфологиялық қызметімен ерекшеленіп тұрады [40]. Автордың айтуынша, бұл етістіктер жақ жалғауларын (I және II жақты) есімдер сияқты бірден қабылдап, нақ осы шақ мағынасын білдіреді және түбір күйінде бұйрық райдың II жағының және нақ осы шақтың III жағының тұлғасында тұра береді. Олар көмекші сөз ретінде қолданыла отырып, көсемше тұлғасында тұрған кез келген етістік тұлғасына көмекші ретінде тіркесіп, нақ осы шақтың мағынасында қызмет атқарады. Бұндай кезде көсемше тұлғалы негізгі етістік өзгеріссіз қалып, жақ жалғаулары өздерінің негізгі мағыналарынан ажыраған *отыр, тұр, жатыр,*

жүр көмекші етістіктеріне жалғанады. Жазушы тілінде аталған етістіктерден өрбіген сөзтұлғалардың жазылуына осы етістіктердің осы жұмсалымдық қасиеттері себеп болған деуге болады.

Жиі қолданылатын бірінші жүздіктің ішіндегі етістіктердің қатарын грамматикалық көрсеткіштеріне қарай төмендегідей топтауға болады. Олар – ***-тын*** есімшесі *(болатын)*; ***-ған//-ген*** есімшесі *(келген, болған, алған, қалған)*; ***-ып, -іп, -п*** көсемшесі *(алып, айтып, болып, келіп, беріп, қарап, боп, кеп)*; ***-е*** көсемшесі *(кеше)*; ***-са*** шартты рай тұлғасы *(болса, қалса, алса, берсе, келсе, айтса, кетсе)*.

Романдағы етістік тұлғалардың жиі қолданылатынына қарап, жазушы аталған сөзтұлғалардың көмегімен және көсемше оралымдар, шартты бағыныңқылы құрмалас сөйлемдер сондай-ақ ашық райлы баяндауышпен аяқталатын басқа да ерекше сипатталатын тұтастықтар құрастырған деген тұжырымдар жасауға болады. Сондықтан ашық райдың септелу қатары көбіне баяндауыштың көрсеткіші ретінде қабылданады.

Жиі қолданылатын сөзтұлғалардың қатарына *бар, жоқ, аз, көп, керек* секілді модаль сөздерді де жатқызуға болады. Заттың нақты мөлшерін, сапасын білдірмесе де, бұл сөздер тілде есім сөздердің де, етістіктердің де қызметін атқарады. Синтаксистік тұрғыдан алғанда олар сөйлемнің толық мүшесі ретінде қолданылады: *кітап жоқ, журнал бар, білім көп, ақысы аз, сия керек*. Бұнда олар өздері жеке тұрып та, тіркесіп келіп те күрделі баяндауыштың қызметін атқара береді: *мен айтқан жоқпын, естуім бар, көп болса, білсең керек* және т.б. Соңғы мысалдарда модаль сөздер басыңқы сыңарды мағыналық жағынан толықтырып, нақтылап тұрады. Осындай жұмсалымдығына орай осы лексикалық топ басқа да лексика-семантикалық құрылымдармен қатар тілде өте жиі қолданылады.

Аса жиі қолданылатын сөзтұлғалардың қатарына есімдіктің түрлерін жатқызуға болады: жіктеу есімдіктері: *мен, сен, ол, оның;* сілтеу есімдіктері: *мынау, осы, бұл, сол;* өздік есімдіктері: *өз, өзі, өзінің,* сұрау есімдігі - *не?* белгісіздік есімдігі - *не*. Олардың жиі қолданылуын түсіндірудің өзі артық. Өйткені кез келген есімдік тілде қолданушының мақсатына қарай есімнің барлық түрінің, сондай-ақ етістіктен болатын есімдердің орнына қолданылады. Есімдіктер мағыналық жан-жақтылығына қарай сөйлемнің тұрлаулы және тұрлаусыз мүшесі де бола алады. Бұның өзі романдағы олардың жиі қолданысын анықтаған біздің статистикалық мәліметтеріміздің дұрыстығына дәлел бола алады.

Үстеулердің ішінен аса жиі қолданылатындардың қатарына негізінен "қашан?" деген сұраққа жауап беретін *енді, қазір, бүгін, әлі* сияқты мезгіл пысықтауыш қызметіндегі үстеулер мен "қандай?" деген сұраққа жауап беретін *ең, дәл* сияқты анықтауыш қызметіндегі үстеулер кіреді. Романның ішкі мазмұнында уақыт факторы маңызды рөл атқарады, сондықтан оның реңктік мәндерінің кейде айқын, кейде бәсең көрінуі заттың сипатын, қимылдың немесе белгілі бір құбылыстың мөлшері мен көлемін білдіруге жоғарыда келтірілген үстеулердің жазушының құралы ретінде жиі қолданылуы көркем шығарманың эстетикалық талаптарынан туындаған.

Аса жиі қолданылатын сөзтұлғалардың қатарына сын есімдер де кіреді. Сапалық – *үлкен, қатты, қалың, жақсы, қара, жас; жаңа, соң* түбірлерінен жасалған туынды сын есімдер *жаңағы, соңғы*. Тілде белгілі бір статистикалық бояуы жоқ болса да бұлардың барлығы да әсіресе сапалы сын есімдер заттар мен құбылыстардың сипатын білдіретіндігіне байланысты шығармада жиі қолданылады. Яғни бұл сөзтұлғаларды жазушы өмірдің бояуын, қоғамдағы құбылыстарды, табиғаттың тамаша сырларын бейнелеу үшін қажетінше жиі пайдаланған. Ал, жалпы, сын есімдердің өзі қолданылуына қарай басқа жиілік аймақтарда да кездесіп отырады.

Біз сипаттап отырған аймаққа зат есімдердің ішінен үш түрлі сөзтұлғалар енеді. Біріншісіне романның басты кейіпкерлерінің есімдері жатады – *Абай, Құнанбай, Базаралы, Әбіш, Ербол.* Тағы бір айта кететін жайт кейіпкерлердің Абайдан өзге барлығының аты тізімде септелмей берілген. *Абай* есімі атау тұлғада келумен қатар ілік *(Абайдың)* және барыс *(Абайға)* тұлғаларында келеді.

Зат есімдердің екінші түріне нақты заттардың атауын білдіретін сөздер жатады: *ел, кісі, үйі* және абстрактілі зат есімдер: *сөз, құн.* Бұл сөздердің барлығы қазақ тілінің сөздік қорының актив қорына жатады. Тек *құн* сөзі ғана қазір көбіне *өзіндік құн* тіркесінде қолданылады. Осы сөздер бейтарап стильдің лексикасы ретінде шығарманың сөздік құрамын жасауда басты рөл атқарған деуге болады.

Зат есімдердің үшінші түріне *кезде, ішінде* сөздері кіреді. Олар мағынасына қарағанда үстеулерге жақын келгенімен, біз оларды *кез* және *іш* түбірлеріне тәуелдік жалғаулары *і – іш+і* мен табыс септігі ***-де/-нде*** *(кезде, ішінде)* жалғанып тұрғандықтан зат есімге жатқызамыз. Бұл сөздердің осы жолмен грамматикализациялануы тілде көмекші сөздер сияқты бұлардың да жиі қолданылуына мүмкіндік жасайды.

Сан есімдердің ішінен жиі қолданылатын тізімге *бір, екі, жалғыз* сөзтұлғалары енді. *"Бір"* сөзі қазақ тілінде болсын, жалпы түркі тілінде болсын өзінің бастапқы мағынасын толық та, жартылай да сақтай отырып өте жиі жұмсалынады. Мәселен, *бір адам* деген тіркес адамның *біреу* екендігін және белгісіз *біреу* екендігін білдіреді. Оның қай мағынада екендігін мәнмәтіннен ажыратуға болады. Бұл сөздің осындай лексикалық әмбебап табиғаты оны әртүрлі стильдік мақсаттарда пайдалана беруге мүмкіндік береді. Бірақ *"бір"* сөзінің осы көп мағынасының ішінен статистикалық санауда негізінен сан есім мағынасы есепке алынды.

"Екі" сөзінің *"бір"* сөзі сияқты бірнеше мағынасы жоқ, сондықтан ол тілде тек сан есім мағынасында жұмсалады. Оның романдағы жиі қолданылуын автордың салыстыру, теңдестіру, қарсы қою және т.б. қолдану барысындағы шығармашылық тәсілі деп түсіндіруге болады.

Жалғызбастылықты білдіретін *жалғыз* және *жалқы* сөздерінің эмоционалдық бояуы қанық. Осындай қолданыстарға лайық сөз болғандықтан оның жиі ұшырауы заңды құбылыс деуге болады.

Алынған нәтижелерге байланысты мынадай сұрау туындауы мүмкін: романның әртүрлі кітаптарында сөздердің жиілігі біркелкі болмаса, оның қосынды түріндегі жиілік сөздігіндегі сөздердің жиіліктік сипатының нақтылы мен дәлдігі қаншалықты? Бірақ жеке кітаптардағы мәліметтерді салыстырғанда, автор өз шығармасындағы стильдік колоритті аяғына шейін сақтағандығын байқаймыз. Жиі қолданылған лексикалық топтар әр кітапта бірдей екендігіне көз жеткізуге болады. Қолдағы мәліметтерге қарағанда 110 сөзтұлғаның романның әр бөлігіндегі қолданысы бірдей. Оларға төмендегідей етістіктер: *емес, болған, отырып, айтып, түсіп, алған, айтқан, алмай, барып, салып, кеткен, көріп, тартып, екенін, берген, барады, болсын, болар, айтады, сөйлеп, көрген, басып, бастап, бере, қалып, қойып, туған, біліп, салды, күле, тауып, айт, кетеді, ертіп, көтеріп, жетіп, десе, атып, жөнелді, кіріп, жеткен, сұрап, білген;* зат есімдер: *ел, күн, Абай, алды, жол, жерде, қасы, үсті, сөзі, көз, көзі, басын, үсті, сыр, ақын, Абайдан, сәт, өңгіме, елге, үйлер, ара, жақ, жел, қолын, бетін, қолына, қан;* есімдіктер: *бұл, өзі, мынау, бірі, соны, өзін, өзіне, кім, міне, оған, одан, қай, қайда, сені, осымен;* сын есімдер: *үлкен, рас, қара, ұзақ, басқа, сондай, бөлек, қазіргі, аппақ, жаман, ашық, артық;* үстеу: *бұрын, тез, әдейі, ақырын, қатар;* көмекші есімдер: *тағы, ме, қана, сайын;* модаль сөздер

керек. Әртүрлі қолданылатын сөздердің тізіміне қолдану жиілігі орташа аймақтан да кіргізілді. Романның бөлігіндегі әртүрлі қолданыстар статистикалық мәліметтерге қарағанда онша көп емес. Мәселен аса жиі қолданылатын 100 сөзтұлғаның ішінен екінші кітаптан мына зат есімдер кірмей қалды: *ән,* кейіпкерлердің аты *Әйгерім, тек* (шектеу мағынасындағы) модаль сөзі, *жігіт* зат есімі, *бүгін* үстеуі және т.б.

Үшінші кітапта жиі қолданылатын 100 сөздің қатарына *Тәжежан, Дәрмен, Оразбай, Мағаш, Әзімбай, Оспан* сияқты кейіпкерлердің, халық атауы қазақ сөзі және көмекші сөздердің фонетикалық варианттары *(-та//-те)* және т.б. Алдыңғы кітаптардағы берілген тізімде бұрын кездеспеген *Дәрмен, Оразбай, Мағаш* деген есімдер аса жиі қолданылатын тізімге төртінші кітап бойынша енді. Романның соңғы кітабы бойынша басқа сөздерден алдыңғы еркін тізімге кірмеген *әсіресе, өзгеше* үстеулері мен *әр* есімдігі енді.

Сонымен, бұл келтірілген мысалдар жиілік сөздіктерді жасаудың қажеттігін, олардан алынған статистикалық сипаттамалардың теориялық және қолданбалы тіл білімінде қаншалықты құндылығын көрсетумен қатар, көптеген болжамдар береді. Бұл болжамдардың құнды болуы әртүрлі тілдерден, жеке автор шығармаларынан, белгілі стильдегі мәтіндерден т.б. жиілік сөздіктерді жасаумен және оларды ықтималды-статистикалық тәсілдермен зерттеуге де байланысты.

1.9. Жиілік сөздіктерді компьютер арқылы алудың біріккен және іріленген алгоритмі

Статистикалық лингвистика саласында қарастырылатын мәселелер, көбіне, мәтіннің әртүрлі бірліктерінің қолдану жиілігін білуді қажет етеді. Ол бірліктер – әріп, әріп тіркесі, буын, буын тіркесі, сөз, сөз тіркесі, сөзтұлға, сөзтұлғалар тіркесі және т.б. болуы мүмкін. Бірліктердің мәтіндегі қолдану жиілігін анықтау үшін ең алдымен сол мәтін бойынша жиілік сөздіктер түзіліп алынуы керек. Қарапайым түрде айтқанда, жиілік сөздік дегеніміз, ол қолдану жиілігі көрсетілген тілдік бірліктер тізімі. Жиілік сөздіктегі бірліктердің, яғни сөз не сөзтұлғалардың қалай сұрыпталуына байланысты олардың типтері ажыратылады. Алдыңғы тақырыпшаларда сөздіктер типтеріне (түрлеріне) сипаттамалар берілгенін еске түсірсек, олардың айырым белгілерін білу қиынға түспейді.

Бұл жұмысты "қолмен", яғни бөлек-бөлек қағаз парақтарына жазып, бірдей тілдік бірліктерді бөлектеп сұрыптап, олардың қайталану санын – жиілігін анықтауға және белгілі бір сипатта сұрыптауға да болар еді. Әрине, жиілік сөздікті "қолмен" түзу көп уақытты, күшті және аса ұқыптылықты қажет етеді. Көп жағдайда, ондай сөздіктердің нәтижесі зерттеушіні қанағаттандыра да бермейді. Сондықтан әрі жиі қайталанатын, әрі көп көлемді, әрі есептеуді және әртүрлі сұрыптауларды қажет ететін бірыңғай операцияға жататын міндеттерді (есептерді) адам көмекшісіне – компьютерге жүктеген жөн. Ол үшін жиілік сөздіктерді "қолмен" жасаудағы ретін сақтай отыра, тізбектелген амал-әрекеттің әр қадамы санадан тыс қалмайтындай етіп, "алгоритм" құрастырылады. Осы алгоритм негізінде компьютерге арналған арнайы "жасанды тіл" көмегімен маман-бағдарламашы (программист) компьютерлік бағдарлама жазады. Бағдарламаның дұрыс-бұрыстығы тиянақты түрде тексеруден өткеннен кейін ғана (отладка) ол іске қосылады. Қазіргі компьютерлер аталған жиілік сөздіктерді әрі тез, әрі қатесіз, әрі зерттеушінің қалауына сай көрнекі түрде орындай алады.

Оқырманды "алгоритм" ұғымымен және оны құрастыру мәселелерімен қарапайым түрде хабардар ету мақсатымен әліпби-жиілік сөздік, жиілік сөздік, кері әліпби-жиілік сөздік және сөзнұсқағыш әліпби-жиілік сөздіктердің біріктірілген және іріленген алгоритмінің көрінісін төмендегі "Сызба-топтама" (блок-схема) арқылы беруді жөн санадық. Әрине, бұл төрт түрлі сөздіктің әрбіреуіне бөлек-бөлек "алгоритмдер" сызба-топтама да жасауға болады. Біздің біріктірілген түрде ұсынуымыз әмбебаптық сипатты қалауды және қысқаша баяндауды көздеуден туындап отыр. Әр сөздіктің арнайы "кілт" (арнайы белгі) арқылы айырым табатынын ескерте отыра, олардың өркайсысының және ортақ тұстарының сипаттамаларына тоқталайық.

Сызбада көрсетілген "А0" топтамада алдын ала келісілген сөздік типтерін ажыратуға қажетті "кілт" шамалары белгіленеді. Егер бағдарлама жұмысы әліпби-жиілік сөздік алуға бағытталған болса, "кілт" мәні *K=1*, ал жиілік сөздік үшін *K=2*, кері-әліпбилі жиілік сөздік бойынша *K=3*, сол сияқты сөзнұсқағыш әліпби-жиілік сөздік түзу кезінде "кілт" мәні *K=4* деп шартты түрде қабылданды. Ал "кілт" мәнінің осы төрт түрлі санның біреуіне тең болу шарты бұзылса – қателік, яғни күтілмеген жағдай (авария) деп саналады да бағдарлама жұмысы тоқтатылады.

Сөздік алынатын мәтін бір немесе бірнеше бөліктерден тұруы да мүмкін, сондықтан компьютер жадына енгізілуге тиісті бөліктер санын қадағалау сызбаның "А1" топтамасында іске асады. Мәтін бөліктері *"i"* мәні бойынша реттеледі.

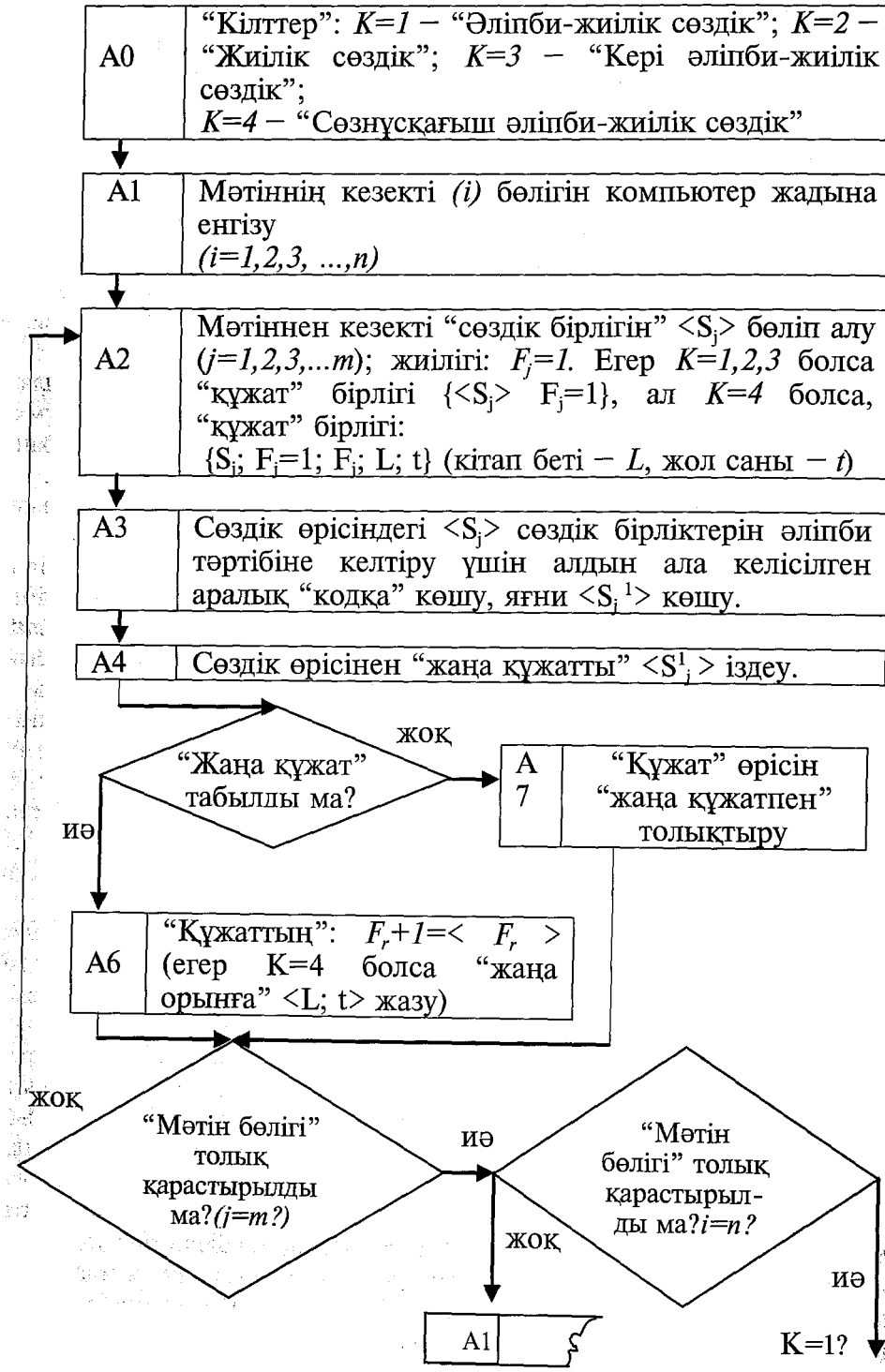
Мысалы, мәтін 5 бөліктен тұрса *(n=5), i=1, i=2, i=3, i=4, i= n=5* деп компьютер жадына өз кезегімен енгізіледі.

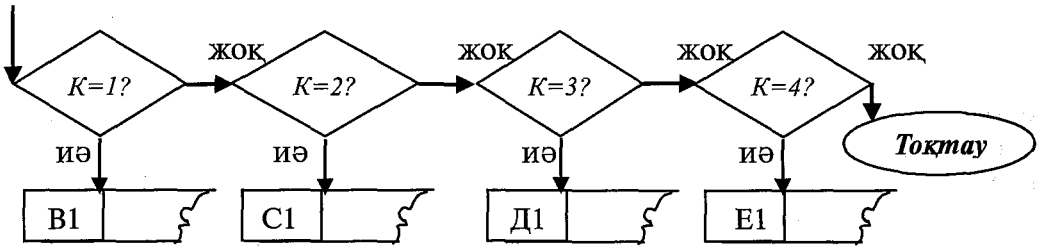
Сызбаның "А2" топтамасында мәтін бойынан кезекті сөздік бірлігін бөліп алу процесі орындалады. Мысалы, сөздік бірлігі сөз не сөзтұлға болса, оның мәтіндегі шекаралық формальды белгілері: "бос орын" (пробел), тыныс белгілері, тырнақша және т.б. деп саналады. Ал қажетті деген мәтін бірлігі осындай белгілердің арасында орналасқан әріптер тізбегі ретінде формальды түрде бөлініп алынады. Егер мәтін бөлігіндегі сөзтұлға санын *"m"* деп белгілесек, *j=1* болғанда бірінші сөзтұлға S_1, ал *j=2* болса, S_2 және осылайша *j* мәні өсіп барып, *j=m* болғанда S_m болады, яғни бұл жағдайда мәтін бөлігіндегі сөзтұлғалар толығымен қарастырылды деген сөз. "Кілт" мәніне қарай, осы топтамада "құжат" бірлігі (жиілік сөздік бірлігі) қалыптасады. Егер, "кілт" мәні *K=1, 2, 3* болса, "құжат" бірлігі $\{S_j; F_j=1\}$, ал *K=4* болса – $\{S_j; F_j=1; L; t\}$ түрінде қалыптасады (*"L"* – кітап беті, *"t"* – беттегі жол саны).

Бүгінгі таңда компьютерлер орыс және латын әліпбилеріне ғана бейімделген, ал қазақ әріптерінің әліпби тәртібімен орналасуына әлі де толық мүмкіншілік жоқ. Осы себептен, қазақ әріптерін әліпбилеуге мүмкіндік жасайтын "аралық кодқа" уақытша көшуге мәжбүр боламыз. Сондықтан сызбаның "А3" топтамасында әрбір "құжат" бірлігіндегі сөзтұлғаны (не басқа бірлікті) аралық "кодқа" ауыстыру жүргізіледі, яғни $<S^1_j>$-ге ауыстыру іске асады. Енді "А4" топтамасының атқаратын қызметі, ол жиілігі бірге тең және аралық "кодта" жазылған "құжатты" осыған дейінгі түзілген (жазылған) сөздік өрісі ішінен іздеу. Нәтижесінде – не табылды, не табылмады деген жауап күтіледі. Осы жауаптар бойынша әрі қарайғы бағдарлама жұмысының бағыты "А5" топтамасында қарастырылады. Егер түзіліп жатқан сөздікте ондай бірлік табылса, онда "А6" топтамасында ол бірліктің жиілігіне "1" саны қосылады. Ал сөздік өрісінен іздеген құжатымыз табылмаған жағдайда "А7" топтамасында сөз жиілігі "1-ге" тең жаңа құжатпен толықтырылады. Бұл екі жағдайдан кейін, яғни "А6" және "А7" топтамалар жұмыстарынан кейін, әрі қарайғы

1-сызба-топтама

Жиілік сөздіктерді компьютер арқылы алудың біріккен және іріленген алгоритмінің сызба-топтамасы





С1	Сөздік өрісіндегі "құжат" жиілігінің *(F_r-дің)* кему тәртібімен сұрыптау, ал $F_r=F_{r+1}$ болса, ондай сөздік бірліктерін әліпби тәртібімен сұрыптау және бастапқы "кодқа" ауыстыру
D1	Сөздік өрісіндегі "құжат" бірлігін $< S_r >$ "теріс ретте" жазу (реверстеу), әліпби тәртібімен сұрыптау, бастапқы қалыпқа (ретке) және "кодқа" ауыстыру
Е1	Сөздік өрісіндегі "құжат" бірлігін S_r әліпби тәртібімен сұрыптау және бастапқы "кодқа" ауыстыру. Сөзнұсқағыш сөздікті қалыптастыру: r – рет саны; S_r – сөздік бірлігі; $(L; t)$ – орны; F_r – жиілігі
В1	Сөздік өрісіндегі "құжат" S_r бірлігін әліпби тәртібімен сұрыптау және бастапқы "кодқа" ауыстыру
В2	Сөздікті қалыптастыру: r – рет саны; S_r – сөздік бірлігі, F_r; $\sum_{r=1}^{n} F_r$; $f_r = \frac{F_r}{N}$; $\sum_{r=1}^{n} f_r$.
В3	Сөздікті қағазға шығару және компьютер жадындағы "сөздік қорына" қосу

Соңы

бағыт "А8" топтамасында түйіседі. Бұл жерде мәтін бөліктері толығымен қарастырылып болу-болмауы тексеріледі (яғни $j \leq m$ болуы). Егер мәтін бөлігі толығымен қарастырылып болмаса, бағыт "А2" топтамасына қайтып оралып, жоғарыда аталған жұмыс түрлері (ішкі цикл ретінде) қайталанып, мәтін бөлігі толық аяқталғанға дейін жалғасады. Бұдан әрі қарайғы жұмыс бағыты екіге тармақталады. Біріншісі – мәтін бөлігі толық қарастырылып болған жағдай. Бұл бағыт бойынша бағдарлама жұмысы "А1" топтамасына барып, кезекті мәтін бөлігінен сөздік бірліктерін бөліп алу процесіне сыртқы цикл ретінде қайта оралады. Екінші бағыт, ол мәтін толық қарастырылып болған жағдайға қатысты *(i=n)*. Әрі қарайға жұмыс "кілттің" мәніне қатысты бағытталады.

1. Егер "кілт" мәні *K=1* болса, ол әліпби-жиілік сөздік түзуге байланысты жолмен жүруі қажет. Сондықтан түзілген сөздікті "В1" топтамасында әліпби тәртібімен сұрыптау іске асады да, сосын сөздік бірлігі (сөз не сөзтұлға) алғашқы "код" қалпына ауыстырылады. Ең соңында, яғни "В2" топтамасында қажетті деген сөздік пішіні қалыптастырылып, "В3"-те, зерттеушінің қалауы бойынша, нәтиже қағаз бетіне шығарылады және компьютерлік сөздік қорына қосылады. Соңғы нәтиже компьютер жадында орналасып, әліпби-жиілік сөздікпен (не сөздіктің басқа түрімен) зерттеуші-тілшінің пайдалануына көп жеңілдік келтіретіні айқын.

2. Егер "кілт" *K=2* болса, "С1" топтамасында бағдарлама "жиілік сөздік" түзу үшін, сөздік бірлігінің абсолютті жиілігінің кему тәртібі бойынша, ал бірдей жиілікті бірліктерді (сөз, сөзтұлғаларды) әліпби ретімен сұрыптауды іске асырғаннан кейін сөздік бірлігін алғашқы "код" мәніне ауыстырады. Осыдан кейін, бағдарлама жұмысы және оның қызметі жоғарыда айтылғандардан мәлім, "В2" топтамасына бағышталады.

3. Егер "кілт" *K=3* болса, жұмыс бағыты "D1" топтамасына бағытталады. Ал мұнда "кері әліпби-жиілік сөздік" алу мақсатымен сөздік өрісіндегі "сөз" не "сөзтұлғалар" теріс қарай қайта жазылады – реверстенеді. Бұл әрекет бірліктерді соңғы әріптері жағынан әліпби тәртібімен реттеуге мүмкіндік береді. Әліпби тәртібіне келтіру процесі аяқталғаннан соң, сөздік бірліктері алғашқы қалпына келтіріліп, бастапқы "кодына" қайта ауыстырылады. Ендігі жұмыс түрі жоғарыда айтылғандардан мәлім, ол "В2" және "В3" топтамаларында атқарылатын соңғы реттегі бағдарлама жұмысы.

4. "Кілт" *K=4* жағдайында әрі қарайғы бағдарлама жұмысы "Е1" топтамасынан жалғас табады. Мұндағы жұмыс түрі "В1" топтамасындағыны қайталайды, тек одан айырмашылығы "Е1" топтамасында "сөзнұсқағыш әліпби-жиілік сөздік" пішіні қалыптасады, сөздік бірліктерінің кітаптағы кездесу жиілігіне сай, олардың орны: кітап беті *(L)* мен жол саны *(t)* да көрініс табады. Аталған "В1" мен "Е1" жұмыс топтамаларындағы қайталаулар сияқты сызбаның "В1", "С1", "D1", "Е1" топтамаларында: жиілік сөздіктерді сұрыптау ісі аяқталғаннан соң, сөз не сөзтұлғаны аралық "кодтан" бастапқы "кодқа" келтіру жұмыстары қайталанып, бірдей әрекеттер жасалады.

Сызба-топтама жұмысын сипаттауды қорыта келе айтпақ ойымыз, осы сияқты "сызба" құрудың басқа да ұтымды жолдары болуы мүмкін [31, 30-40]. Сондықтан сөздік жасауға қажетті компьютерлік бағдарламаның сызба-топтамасы тек осылай ғана сызылу керек екен деген пікір тумауы керек. Ұтымды сызба-топтама жасау шығармашылықты, тапқырлықты, шеберлікті және т.б. қасиеттерді қажет етеді. Себебі әрі қарайғы "сызбаның" "жасанды тілде" көрініс табуы және оның компьютерде ұтымды іске асуы сызба-топтаманың сапалы құрылуына көп байланысты.

Екінші тарау
СТАТИСТИКАЛЫҚ ЛИНГВИСТИКА

2.1. Статистикалық заңдылық және ықтималдық

Біз өмір сүретін ортадағы заңдылықтардың екі түрін арнайы бөліп айтуға болады. Оның біріншісі динамикалық деп аталса, екіншісі статистикалық (ықтималдық) заңдылыққа жатады. Бірінші заңдылық бойынша оқиғаның болу-болмауын алдын ала дәлме-дәл айтуға болады. Мысалы, темірдің суда бататынын, ток қосылғанда электр шамының жарық беретінін, 100°С-да судың қайнайтынын, түннің күнге ауысатынын және т.б.

Ал енді заңдылықтың екінші түрінің нәтижесін алдын ала кесіп айту қиын, оны тек екі аралық арасында, бір *орта шамадан ауытқуын* ескере отырып қана айтуға болады. Бұл заңдылықтың түріне мысал ретінде адамның ми жұмысын, мектептің, үгіт-насихаттың, мәдениеттің адамға әсерін, баланың психикасының дамуын, сөйлеу қызметін, тілдің дамуы мен қызметін және т.т. айтуға болар еді.

Статистикалық заңдылықтың әсерін мынадай қарапайым екі мысалдан анық аңғаруға болады.

Бірінші: *ойынға бейімделген 6 жақты кубті бірнеше рет жоғары лақтыратын болсақ, оның 1-ден 6-ға дейінгі жақтары неше рет түсетінін алдын ала шамамен есептеп шығаруға бола ма?*

Екінші: *екі жақты тиынды бірнеше рет лақтырудан кейін, оның қай жағы неше рет жоғары қарап түсетінін алдын ала шамамен есептеп шығаруға бола ма?*

Сонымен, кубті 600 рет жоғары лақтырдық делік, сонда оның әрбір жағының неше рет жоғары қарап түсуін білу үшін 600 санын 6 санына бөлгенге тең шама болады екен, яғни 600:6=100 рет не болмаса оның шамасы осы 100 санынан не көп, не аз болып ауытқуы мүмкін. Сол сияқты, тиынды 500 рет жоғары лақтырсақ, оның әр жағы 500:2=250 рет шығуы мүмкін немесе осы санға шамалас болады.

Екі мысалда да күтілетін оқиғаның шығуына әртүрлі *кездейсоқ жағдайлар* себепші болады. Мысалы, кубик пен тиынды жасалу пішінінің дұрыс-бұрыстығына, ауа кедергісіне, тәжірибешінің лақтыру күшіне, түсетін жерінің тегістігіне және басқа да көптеген сырт жағдайлардың әсерінен кубтің таңдалған жағы 100 рет, ал тиынның бір жағы 250 рет шығу мүмкіндігінен ауытқиды. *Ауытқу шамасы* кездейсоқтыққа тәуелді болғандықтан, ол ықтималдық заңдылығына бағынады. Мұндай заңдылық математика жолымен, яғни ықтималдық теориясы мен математикалық статистика әдіс-тәсілдері арқылы тағайындалады [17, 29, 30].

Көптеген тәжірибелер жүргізіп, сосын барып ауытқудың орта шамасын есептеуге де болады. Бірақ ықтималдықтың теориясы негізінде тәжірибе жүргізбей-ақ, оқиғаның мүмкіндігін, яғни куб пен тиынды жоғары лақтырған кезде неше рет шығу ықтималдығын (мүмкіндігін) есептеп шығаруға болады.

Айталық, таңдалып алынған кубтің не тиынның кез келген жағының шығуын "оқиға" деп ат қойып, оның атын "*А*" деп белгілесек және "*т*" саны арқылы "оқиғаның" *тәжірибе үстінде* орындалу (қалаған жақтың шығу) санын белгілесек,

"*n*" арқылы тәжірибенің (жоғары лақтыру) жалпы санын, яғни элементар оқиғалар санын белгілесек, онда *оқиғаның ықтималдығы* классикалық анықтама негізінде мына өрнекпен анықталады [17, 16-22]:

$$P(A)=\frac{m}{n} \qquad (1)$$

Бұл теңдіктегі $m \le n$ болғандықтан $P(A) \le 1$.

Егер *күткен оқиға* (тиынның не күбтің таңдалған жағы) тәжірибе үстінде бірде-бір рет шықпаса (пайда болмаса), яғни жағдайдың барлығы да күтетін оқиғамызға "қолайсыз" болса, онда $m=0$ және $P(A)=\frac{0}{n}=0$ болады.

Егер тәжірибенің өн бойы күткен оқиғамыз үнемі шығып отырса, яғни жағдайдың бәрі де күтілетін оқиға үшін қолайлы болса, онда $m=n$ және $P(A)=\frac{m}{n}=\frac{n}{n}=1$ болады.

Айтылғандарды қорыта келе (1) өрнекке қатысты мынандай анықтама беруге болады: егер $P(A)=0$ болса, ондай оқиға – *"мүмкін емес оқиға"* деп, ал $P(A)=1$ болса, *"ақиқат оқиға"* деп аталады.

Сонымен, *ықтимал оқиғаның* барлық жағдайын қамтитын математикалық өрнек түрі, ол $0 \le P(A) \le 1$ деген теңсіздікпен көрініс табады.

Мысалы: *М. Әуезовтің "Абай жолы" романының 2-ші кітабында 124398 сөзқолданыс бар. Оның 11467 сөзқолданысы сын есім сөздер екен. Сонда 4 кітаптан тұратын "Абай жолы" романында сын есімнің кездесуінің ықтималдығы қанша?* – деген сұрақ туындауы мүмкін. Бұл сұраққа $P(A)=\frac{m}{n}$ өрнегі жауап береді: $n=124398$, $m=11467$, A – сын есімнің кездесуін білдіретін оқиға.

Сонда, сын есім сөздердің роман бойында *кездесу ықтималдығы* – $P(x)=\frac{m}{n}=\frac{11467}{124398}=0,092 \approx 0,1$ шамасына тең болады. Бұл шаманың аз не көп екендігін білу үшін "0" мен "1" аралығымен салыстыру керек. Егер $P(x)$ мәні "1" санына неғұрлым жақындаған сайын, ол оқиғаның шығу мүмкіндігі артады.

Оқиғаның ықтималдығын анықтаудың ең қарапайым *статистикалық құралы* ретінде ол оқиғаның кездесу *"жиілігін"*, *"орта жиілігін"* және *"орта жиіліктен ауытқу"* деп аталатын шамаларды табу жолдарын айтуға болады.

Бұл жердегі *"жиілік"* терминін *"болмыстың бөлігі"* ішіндегі байқауға алынған оқиғаның (қайсыбір тілдік бірліктің) кездесу саны деп түсіну қажет. Ал *"болмыстың бөлігі"* ретінде, мысалы, үлкенді-кішілі көлемдегі мәтін бөлігі алынуы мүмкін.

Айталық, ойын кубін *1000* рет жоғары лақтырғанда оның *"бір"* деген белгісі бар жағы *170* рет шықты десек, онда осы сан *(170)* *"бірдің кездесуі"* жайлы оқиғаның *"жиілігі"* деп есептеледі. Сол сияқты, М. Әуезовтің "Абай жолы" романында түбір есімдік *20811* рет, ал оған тәуелдік жалғауы жалғанған түлғасы (формасы) *2686* рет кездеседі делік. Осы келтірілген *20811* бен *2686* сандары есімдік сөздердің роман мәтіні бойынша есептелген *"жиілігі"* болып саналады.

Әдетте, статистиктер тілдік бірліктерге қатысты *оқиғаның заңдылығын* ашу кезінде бір саладағы барлық мәтіндерді түгелімен (*генералды жиын* немесе *бас*

жиын) қамтуды мақсат етпейді, себебі ол мүмкін де емес. Сондықтан зерттеуші ол мәтіннен тек сынаққа түсетін белгілі көлемдегі үлгілерімен ғана қанағаттанады. Міне, осындай мәтін бөліктерін *"таңдама бөліктер"* немесе *"таңдама жиындар"* (выборки) деп атайды. Әрине, мұндай бөліктер көлемі жағынан бас жиыннан айтарлықтай аз мөлшерде болады және бөліктер саны бірнешеу болған жағдайда, олардың көлемдері тең не жуық шамада болуы керек. Осындай *"сынама үлгілер"* бойынша тағайындалған "оқиғаның" статистикалық заңдылығы бас жиынға да таралады. Мәтіннің осындай *сынама үлгілері* (немесе *таңдамалары*) бойынша есептелген "оқиғаның" кездесу жиілігін *"таңдама жиілік"* деп атайды. Мысалы, көркем әдебиет стилі *бас жиын* деп саналса, М.Әуезовтің 20 томдық шығармалар жинағы *таңдама бөлік* болып саналады. Ал егер бас жиын ретінде М.Әуезовтің 20 томдық шығармалар жинағының мәтіндері алынса, онымен салыстырғанда "Абай жолы" романының бір ғана томының мәтіні *"таңдама бөлік"* болып есептеледі. Осы *таңдама бөліктен* анықталған тілдік бірліктің жиілігі, мысалы түбір есімдіктің жиілігі, *таңдама жиілік* деп аталады. *Таңдама жиілік* жеке тұрып "оқиғаның" ықтималдығы мен статистикалық заңдылықтары жайлы зерттеушіге қанағаттандырғысыз ақпарат беруі мүмкін. Ал егер біз *"орта таңдама жиілік"* немесе *"орта жиілік"* деген терминдермен аталатын шаманы анықтайтын болсақ, бұл жағдайда *"оқиға"* туралы ақпарат ұлғая түседі.

Орта таңдама жиілікті есептеудің көптеген жолдары ішінен біз тек қана ең қарапайым түрін қарастырмақпыз [17; 29; 30, 22].

Айталық, зерттеуге жататын мәтін ішінен құрылымы жағынан біркелкі келетін және бірдей көлемдегі (ұзындықтағы) бірнеше *таңдама мәтін үзінділерін* зерттеу нысаны ретінде алдық делік. Мысалы, таңдалып алынған зерттеу нысаны – *500 негіз сөз* тұратын *10 мәтін бөліктері (үзінділері)* болсын. Енді зерттеу *"оқиғасы"* ретінде "зат есімнің кездесуін" *"х"* деп белгілейтін болсақ, оның *"орта жиілігін"* есептеу үшін *10* мәтін үзінділердің әрбіреуіндегі зат есім сөздердің кездесу жиілігін санап шығуымыз керек болады. Егер ондай жиіліктерді $x_1,x_2,...,x_{10}$ деп белгілесек және тәжірибе жүзіндегі олардың сандық мәндері: $x_1=182$, $x_2=187$, $x_3=218$, $x_4=173$, $x_5=158$, $x_6=201$, $x_7=222$, $x_8=233$, $x_9=213$, $x_{10}=194$ шамаларына тең деп есептейік.

Енді *"орта жиілікті"* $\overline{X}$ және мәтін үзінділерінің санын n деп белгілесек, *орта жиілікті* есептеп шығару үшін мынандай амалдарды орындауымыз керек:

а) оқиғаның кездесу жиіліктерінің қосындысын табу:

$$x_1+x_2+...+x_{10}= =182+187+...+194=1981;$$

ә) *"орта жиілік"* шамасын немесе *арифметикалық ортаны* табу:

$$\overline{X}=\frac{x_1+x_2+...+x_{10}}{n}. \qquad (2)$$

Жоғарыда келтірілген есеп берілісі бойынша зат есімдерге қатысты *"орта жиілік"* (2) өрнек арқылы есептеп шығару керек болса, ол мына түрде табылады: $\overline{X}=\frac{1981}{10}=198,1\approx 198.$

Сонымен, алынған сандық нәтиженің мәні мынада: мәтін ішіндегі әрбір 500 негіз сөзге, орта есеппен алғанда, 198 зат есім сәйкес келеді деп саналу қажет.

Егер математика пәнінде қосындыларды мынадай "$\sum$" таңбамен белгіленетінін ескерсек, онда: $x_1+x_2+...+x_{10}=\sum_{i=1}^{10} x_i$ пішінінде жазылады. Онда *"орта жиілікті"* анықтаудың (2) өрнегін қысқа түрде былай жазуға болады:

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i . \qquad (3)$$

Бұл (3) өрнек *"орта таңдама жиілікті"* (немесе *"орта жиілікті"*) анықтау үшін қолданылады.

Есеп: *Жоғарыда қарастырған таңдама мәтіндер ішінен сын есімдердің кездесу жиіліктері:* $x_1=69$, $x_2=71$, $x_3=83$, $x_4=60$, $x_5=43$, $x_6=73$, $x_7=72$, $x_8=59$, $x_9=69$, $x_{10}=71$. *Сын есім сөздердің "орта жиілігін" табу керек. Яғни егер* $n=10$ *және сын есімнің кездесу жиіліктері белгілі болса,* $\overline{X}=?$ (*жауабы:* $\overline{X} \approx 67$).

Тілді статистикалық тәсілмен зерттеуде *"орта жиілік"* шамасын білу аса маңызды деп саналады. Зерттеушінің таңдап алған тілдік *"оқиғасының"* статистикалық заңдылығын айқындау немесе оның ықтималдығын білу осы (3) *"орта жиілік"* өрнегі арқылы іске асады. Сандық мәні белгілі болған *орта жиілікті* әрі қарайғы зерттеуімізге пайдалану үшін *таңдама жиіліктің* (x_i) *орта жиіліктен* ($\overline{X}$) *ауытқу* шамаларын анықтау қажет болады.

Егер *"таңдама жиілік"* мәні *"орта жиіліктен"* кіші болса, ауытқудың таңбасы – теріс (минус), ал керісінше жағдайда – оң (плюс) деп саналады. Мұндай ауытқу шамалары жеке тұрып, зерттеушіге беретін мағлұматы аз болуы мүмкін. Сондықтан ауытқуларды да жалпылайтын *"ортаны"* табудың қажеттігі туындайды. *Математикалық статистика* пәнінде мұндай "*ортаны*" табудың екі түрлі жолы кездеседі:

а) *ауытқулардың абсолют шамаларының ортасын* табу. Бұл жағдайда теріс таңбалы айырым шамалар оң таңбаға ауыстырылып, содан кейін барып ауытқулардың ортасы табылады. Ал ондай ортаны табу үшін абсолютті ауытқулардың қосындысын таңдама бөліктердің санына бөлу қажет:

$$\overline{a} = \frac{\sum / x_i - \overline{x} /}{n}; \qquad (4)$$

ә) немесе екіге дәрежеленген ауытқудың орта шамасын мынандай өрнекпен анықтау қажет:

$$\sigma = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}} . \qquad (5)$$

Мұндағы σ *(сигма)* – *квадраттық ауытқудың ортасы (среднее квадратичное отклонение)* және $x_i - \overline{x}$ – *таңдама жиіліктің орта шамадан ауытқуы* деп аталады.

Егер $x_i - \overline{x} = a_i$ арқылы белгілейтін болсақ, онда *квадраттық ауытқудың ортасы*:

$$\sigma = \sqrt{\frac{\sum a_i^2}{n}} . \qquad (6)$$

Бұл (6) өрнекті (формуланы) сынақ таңдамалардың (бөліктердің) ұзындығы (көлемі) өзара тең болған жағдайда пайдалану қажет (мысалы, *500* сынақ таңдамалардың әрбіреуі *100* негіз сөзден тұрады).

2. Жоғарыдағы (5) теңдіктің екі жағын бірдей квадраттайтын болсақ:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}. \qquad (7)$$

Математикалық статистика пәнінде (7) өрнек *"дисперсия"* деген атпен аталады. Кейбір тәжірибе кезіндегі есеп-қисаптарда "квадраттық ауытқу ортасы" (6) өрнегінен гөрі *дисперсия* өрнегін (7) пайдаланған ыңғайлы деп саналады.

Есеп: *Бір мәтіннен 5 сынақ таңдама бөлік алынған. Әрбір бөлік 500 атаушы сөздерден тұрады. Оларда етістіктердің кездесу жиіліктері мынадай:* $x_1 = 95; x_2 = 87; x_3 = 94; x_4 = 104; x_5 = 100.$

Орта квадраттық ауытқуды табу керек.

Шешімі: *Орта квадраттық ауытқудың* өрнегі:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}; \quad 1 \le i \le 5; \; n = 5;$$

$$\bar{X} = \frac{95 + 87 + 94 + 104 + 100}{5} = \frac{480}{5} = 96;$$

$x_1 - \bar{X} = 95 - 96 = -1;\; x_2 - \bar{X} = 87 - 96 = -9;\; x_3 - \bar{X} = 94 - 96 = -2;$

$x_4 - \bar{X} = 104 - 96 = 8;\; x_5 - \bar{X} = 100 - 94 = 4;\; (x_1 - \bar{X})^2 = 1;$

$(x_2 - \bar{X})^2 = 81;\; (x_3 - \bar{X})^2 = 4;\; (x_4 - \bar{X})^2 = 64;\; (x_5 - \bar{X})^2 = 16;$

$$\sigma = \sqrt{\frac{1 + 81 + 4 + 64 + 16}{5}} = \sqrt{\frac{166}{5}} = \sqrt{33,2} = 5,76.$$

$\sigma = 5,76$ *– орта квадраттық ауытқу.* $\sigma^2 = 33,2$ *– дисперсия.*

$\bar{X} - \sigma \le \bar{X} \le \bar{X} + \sigma;\; 90,24 \le \bar{X} \le 101,76.$

Осы есептің шешімін төмендегідей кесте арқылы көрнекі түрде де беруге болады:

Сынақ таңдамалар	Етістік сөздер		
	x_i	$a_i = x_i - \bar{X}$	$a^2_i = (x_i - \bar{X})^2$
1-ші	*95*	*-1*	*1*
2-ші	*87*	*-9*	*81*
3-ші	*94*	*-2*	*4*
4-ші	*104*	*8*	*64*
5-ші	*100*	*4*	*16*
Қосындысы	*480*	*0*	*166*
	$\bar{X} = 96$		
	$\sigma = 5,76;$	$\sigma^2 = 33,2$	

Есеп: *Алдыңғы есептің берілісі бойынша зат есім сөздердің жиілігі белгілі делік.*

Олар: $x_1 = 199,\ x_2 = 205,\ x_3 = 195,\ x_4 = 201,\ x_5 = 210,\ n = 5.$

$$\sigma = ?\quad \sum_{i=1}^{5} x_i = 1010,\quad \overline{X} = \frac{1010}{5} = 202,\quad \sum_{i=1}^{5} a_i^2 = 132,\quad \sigma = \sqrt{\frac{132}{5}} = \sqrt{26{,}4} = 5{,}14.$$

Осы есептің шығару жолын кесте түрінде де көрсетуге болады.

Сонымен, жоғарыдағы баяндауымызда математикалық статистиканың ең маңызды әрі қарапайым түрдегі алты ұғымымен және олардың аталу терминдерімен таныс болдық.

Олар:

1) ықтималдық; 2) статистикалық заң; 3) таңдама жиілік; 4) орта таңдама жиілік; 5) ауытқулардың абсолютті ортасы; 6) орта квадраттық ауытқу (дисперсия).

Статистикалық әдісті қолданғысы келген тілші-зерттеушіге, алғашқы қадам ретінде, ықтималдықтар теориясы мен математикалық статистиканың осы аталған ұғымдары мен сипатталған қарапайым санау құралдары жеткілікті деуге болады.

2.2. Таңдама жиіліктер айырымдарын статистикалық бағалау

Жазба не сөйлеу тілін статистикалық жолмен зерттеу, әдетте, бірдей көлемдегі таңдама бөліктер мәтіндері негізінде жүргізіледі. Осындай жағдайда зерттеу нысаны ретінде алынған бір ғана "оқиға" (мысалы, қайсыбір тілдік бірлік) әр таңдама бөліктерінде кездесу жиіліктерінің (таңдама жиіліктерінің) әртүрлі шамада болуы сол бірліктің статистикалық заңдылығының ақпараттық деректері бола алады.

Айталық, қазақ тіліндегі ғылыми-техникалық шығармалар мәтінінен *5000* сөзқолданыстан тұратын таңдама сынаққа алынды делік. Осы мәтін көлеміндегі сөзқолданысты теңдей етіп *10-ға* бөлсек, *500* сөзқолданыстан тұратын *10* таңдама мәтін бөліктері шығады, яғни $N=5000,\ n=500,\ k=10.$ Әрбір таңдама бойынша есептелген етістіктің таңдама жиіліктері: $x_1 = 98,\ x_2 = 87,\ x_3 = 102,\ x_4 = 105,\ x_5 = 123,\ x_6 = 108,\ x_7 = 85,\ x_8 = 78,\ x_9 = 110,\ x_{10} = 104;$ Етістік жиілігінің бұл ауытқулары тілші-тәжірибешіге қандай мәлімет беруі мүмкін немесе жағдай етістік сөздердің ғылыми-техникалық шығармалар стилі үшін заңды не заңсыз құбылысқа жата ма?

Егер жиіліктер арасындағы мұндай ауытқулар статистика тұрғысынан заңды болса, оларды кездейсоқтық жағдайға қатысты деуге болады. Ал егерде таңдама жиіліктің орта жиіліктен ауытқуы елеулі сипатқа енсе, онда оларды статистикалық заңдылыққа бағынбау салдарынан немесе етістік сөздердің қолдану ықтималдығының тұрақсыздығынан деп түсіну керек. Себебі, жиілік шамасының құбылуы кездейсоқтық сипатқа ие болмауынан және ауытқу дәрежесінің елеулі болып, бір ғана ықтималдыққа бағынбауы салдарынан болып отыр. Сондықтан бұл оқиға (етістік сөздердің кездесуі) ешбір статистикалық заңдылыққа бағынбайды деп саналады. Тәжірибе кезінде зерттеуші осындай жағдайдың себебін анықтай білуі керек. Яғни оқиғаның (біздің мысалда етістіктің жиіліктері) қолданылуының орта жиіліктен ауытқуы елеулі ме, әлде ол кездейсоқтық жағдайға қатысты ма?

Міне, осындай сауалдың жауабын табу жолдарына тоқталайық.

Аталған мәселенің басын ашу математикалық статистика пәнінің міндеті. Дәлірек айтсақ, ол үшін аталған саланың "χ^2 *критерий*" (оқылуы – *хи квадрат*) деп аталатын статистикалық құралын пайдалануымыз керек. Бұл критерий *"Пирсонның* χ^2 *келісім критерийі"* деп те аталады (үшінші тарауда бұл критерийге тағы да тоқталамыз).

Тәжірибелік мәтіннен алынған таңдама бөліктердің көлемі (жоғарыда баяндалған мысалдағыдай) тең болып келетін жағдайда χ^2 критерийінің өрнегі төмендегідей жазылады [17, 29, 30, 28-36]:

$$\chi^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\bar{x}} . \qquad (1)$$

Осы (1) өрнектегі x_i – таңдама жиіліктер, $\bar{X}$ – орта таңдама жиілік, Σ – қосу белгісі. Егер таңдама жиіліктің орта жиіліктен ауытқуын, яғни $x_i - \bar{X} = a_i$ деп белгілесек, онда χ^2 критерийі мынадай шағын түрде көрініс табады:

$$\chi^2 = \frac{\sum_{i=1}^{n} a_i^2}{\bar{X}} . \qquad (2)$$

Өрнектің оқылуы: *хи квадрат* – таңдама жиіліктің орта жиіліктен ауытқу квадраттарының қосындысын орта жиілікке бөлгенге тең.

Ықтималдықтың бір ғана тұрақты мәніне байланысты χ^2 шамасы *"жиіліктің үлестірілу заңдылығына"* бағынады. Яғни ықтималдықтың бір ғана тұрақты мәніне қатысты χ^2 *(хи квадрат)* шамасының әртүрлі жиі, сирек не тіпті сирек болып кездесу мәндері сәйкес келуі мүмкін екен. Осыны ескере отырып, математика мамандары *"хи квадрат"* мәндерінің теория бойынша мүмкін шамаларын анықтап, осы саладағы тиісті әдебиеттерде олар кесте деректері түрінде беріледі. Осы деректер арқылы зерттеуші таңдама жиіліктің орта жиіліктен ауытқуын бағалап, оның статистикалық заңға бағыну не бағынбау дәрежесін анықтай алады. Сондықтан да χ^2 критерий, кейде "χ^2 келісім критерийі" деп те аталады. Әрине, бұл жерде "немен" келісу керек екенін түсініп алу қажет. Енді осы мәселеге толығырақ тоқталайық.

Бұл сауалды тәжірибе жүзіндегі деректерге байланысты есептеп шығарылған χ^2 шамасын, оған сәйкестікте тұратын арнайы кестеде келтірілген χ^2-тың теориялық мәнімен салыстыра отырып, келісім жасалады деп түсіну қажет.

Таңдама жиіліктің орта жиіліктен ауытқуының әртүрлі мәндеріне байланысты χ^2-тың тиісті шамалары *2.1-кестеде* көрсетіледі. Аталған *2.1-кестенің* χ^2-тың деректерін, яғни теориялық мәнін түсіндіру үшін жоғарыда келтірілген ғылыми-техникалық шығармалар мәтініндегі етістік сөздердің кездесу жиілігі бойынша қарастырайық: *(N=5000, n=500, k=10):* $x_1=98$, $x_2=87$, $x_3=102$, $x_4=105$, $x_5=123$, $x_6=108$, $x_7=85$, $x_8=78$, $x_9=110$, $x_{10}=104$;

Орта таңдама жиілікті анықтау өрнегі:

$$\bar{X} = \frac{x_1 + x_2 + \ldots + x_{10}}{n}; \quad \bar{X} = \frac{98 + 87 + \ldots + 104}{10} = \frac{1000}{10} = 100;$$

Енді (1) өрнек бойынша: $\chi^2 = \frac{(-2)^2 + (-13)^2 + \ldots + 4^2}{100} =$

$$= \frac{4 + 169 + 4 + 25 + 529 + 64 + 225 + 484 + 100 + 16}{100} = \frac{1620}{100} = 16{,}2.$$

Сонымен, есептеп шығарылған χ^2-тың тәжірибелік мәні 16,2 тең болды. Кестедегі бірінші "тік бағанадағы" сандар – *"число степеней свободы"*

немесе қазақша *"еркіндік дәреже саны"*, – деп аталады *(А.Ж.)*. Оның мәні таңдама мәтін бөлігінің санынан *(k)* бірді кеміткенге тең болады. Егер *еркіндік дәреже санын* – "*v*" деп белгілесек, онда $v = k-1=10-1=9$;

Ал тік бағанадан таңдалып алынған $v=9$ санына сәйкес жатық жолдағы сандар ішіндегі χ^2 теориялық мәнімен тәжірибелік χ^2 мәні салыстырылады.

Нәтижесінде, тәжірибелік χ^2-қа тең не одан үлкендеу шамасы белгіленіп, сол бағана бойындағы (жоғарыдағы), ықтималдық таңдалады.

2.1-кесте

χ^2 критерийінің теориялық мәні

Еркіндік дәреже саны	Ықтималдықтың үлкен шамасына сай келетін χ^2 (хи квадраттың) теориялық мәні					
	0,95 (95%)	0,75 (75%)	0,50 (50%)	0,25 (25%)	0,10 (10%)	0,05 (5%)
1	-	0,10	0,45	1,32	2,71	3,84
2	0,10	0,58	1,39	2,77	4,61	5,99
3	0,35	1,21	2,37	4,11	6,25	7,81
4	0,71	1,92	3,36	5,39	7,78	9,49
5	1,15	2,67	4,35	6,63	9,24	11,07
6	1,64	3,45	5,35	7,84	10,64	12,59
7	2,17	4,25	6,35	9,04	12,02	14,07
8	2,73	5,07	7,34	10,22	13,36	15,51
9	3,33	5,90	8,34	11,39	14,68	16,92
10	3,94	6,74	9,34	12,55	15,99	18,31
14	6,57	10,17	13,34	17,12	21,06	23,68
15	7,26	11,04	14,34	18,25	22,31	25,00
19	10,12	14,56	18,34	22,72	27,20	30,14
20	10,85	15,45	19,34	23,83	28,41	31,41
24	13,85	19,04	23,34	28,24	33,20	36,42
25	14,61	19,94	24,34	29,34	34,38	37,65
29	17,71	23,57	28,34	33,71	39,09	42,56
30	18,49	24,48	29,34	34,80	40,26	43,77

Еркіндік дәреже саны $v=9$ мәніне және тәжірибелік $\chi^2=16,2$ мәніне шамалас кестедегі теориялық χ^2-тың *16,92* мәні және ықтималдықтың *0,05 (5%)* сай келеді екен. Бұл жерде "хи квадрат" мәнінің тәжірибелік шамасы теориялықтан кіші екендігін ескере отырып, былай пайымдауға болады: оқиғаның шығуының барлық 100 мүмкіндігінің тек 5 жағдайында ғана етістік жиіліктерінің жоғарыда көрсетілгендей ауытқу шамаларының кездесуі мүмкін. Сондықтан ондай ауытқуларды *кездейсоқ оқиғаға* жатқызып, етістіктің таңдама жиіліктері статистикалық заңға қайшы келмейді және етістіктің 100-ге тең орта жиілігі 500 сөзқолданыстан тұратын мәтінге тән жиілік деуге болады. Ал етістік сөздердің мәтінде кездесу ықтималдығы $P(x)=\frac{100}{500}=0,2$ шамасына тең. Сонымен, бұл есептің нәтижелері бойынша тәжірибелік ғылыми-техникалық шығармалар мәтінінің барлық сөзқолданысының 20 пайызы етістік сөздерден тұрады деп қорытындылауға болады.

Статистик мамандар таңдама жиіліктің орта жиіліктен ауытқуының *елеулі* не *елеулі еместік* сипатын анықтау үшін χ^2 мәнін кесте бойынша екі ықтималдық шамаларының аралығында қарастырады, яғни *0,95-тен 0,05* аралығы. Осы аралықтан алынатын χ^2 мәндері оқиға ықтималдығының тұрақтылығын және ауытқудың елеулі еместігін білдіреді.

Зерттеуші мамандардың тұжырымдауынша, χ^2 критерийін таңдама жиіліктерінің шамасы 20-дан жоғары және тіпті жүздеп саналған жағдайларда қолданғанды дұрыс деп табады.

Қорыта айтқанда, χ^2 критерийін мынандай мазмұндағы есептерге пайдаланған жөн:

1) жоғарыда қарастырғандай, бір типті мәтіннен бірдей көлемдегі таңдама мәтіндер алынған жағдайда;

Мысалы, есеп берілісі мынадай: $n=500$, $k=5$, сын есім жиіліктері: $x_1=70$, $x_2=82$, $x_3=68$, $x_4=80$, $x_5=75$. $\overline{X}=?$ $\chi^2-?$

Егер $v=k-1=5-1=4$ болса, *1-кестені* пайдаланып, ауытқудың елеулі сипатта еместігін дәлелдеу керек (жауабы $-\chi^2=1,97$).

2) екі түрлі (типті) жиыннан (қатаң түрде) бірдей көлемдегі таңдама мәтіндер бойынша есептелген бір ғана "оқиғаның" (мысалы, зат есімдер) жиіліктерінің ауытқуы кездейсоқ екендігін немесе ауытқу дәрежесінің елеулілігін анықтауда.

Мәселен, есеп берілісі мынадай делік: бірдей көлемдегі *(L=1000)*, стильдік ерекшеліктері бар екі түрлі мәтін таңдамасы беріледі. Бұл мәтін бөліктерінде зат есім сөздердің кездесу жиілігі: $x_1=270$ және $x_2=220$.

Жиіліктердің орта жиіліктен ауытқуының елеулі не елеулі еместігін анықтау керек.

Шешімі: Арифметикалық орта: $\overline{X}=\frac{270+220}{2}=245.$

Ал χ^2-тың тәжірибелік мәні:

$$\chi^2=\frac{(x_1-\overline{X})^2+(x_2-\overline{X})^2}{\overline{X}}=\frac{(270-245)^2+(220-245)^2}{245}=5,1.$$

Енді осы χ^2-тың тәжірибелік мәнін оның кестедегі теориялық мәнімен салыстыра бағалау қажет.

Ол үшін алдымен *еркіндік дәреже санын* анықтаймыз ($v=2-1=1$), ал сосын барып, χ^2-тың кестедегі теориялық мәнінің *3,84*-ке тең екендігіне көз жеткіземіз.

Енді χ^2 пен v мәндерінің негізінде кесте деректері бойынша $\chi^2=3,84$ екендігін және оның ықтималдық пайызы өте аз дәрежеде – *5%*, ал $5,1>3,84$ болғандықтан ауытқу елеулі сипатта және оқиға кездейсоқ емес деп қорытындылаймыз.

3) Салыстыра зерттейтін мәтіндердің көлемі тең болмаған жағдайға байланысты.

Мысалы: *Екі түрлі мәтіндер көлемдері: $L_1=530$, $L_2=970$ және әрбір мәтіндегі сын есім жиіліктері: $x_1=75$ және $x_2=100$.*

Таңдама жиіліктердің орта жиіліктен ауытқу шамасы елеулі ме әлде ол кездейсоқтық сипатта ма?

Шешімі:

а) $x_1+x_2=75+100=175$; *ә) $L=L_1+L_2=530+970=1500$;*

б) $p(x)=\frac{175}{1500}=0,116$; *в)* $\overline{X_1}=\frac{x_1+x_2}{L_1+L_2}\cdot L_1=0,116 \text{ x } 530=61,5$;

г) $\overline{X}_2=\frac{x_1+x_2}{L_1+L_2}\cdot L_2=0,116 \text{ x } 970=113,5.$

Енді $\chi^2=\frac{(x_1-\overline{x}_1)^2}{\overline{x}_1}+\frac{(x_2-x_2)^2}{\overline{x}_2}=\frac{(75-61,5)^2}{61,5}+\frac{(100-113,5)^2}{113,5}=4,57.$

Еркіндік дәреже саны: l=2-1=1; Хи квадраттың кестелік (теориялық) мәні: $\chi_k^2=3,84$ және оның тәжірибелік мәні: $\chi_t^2=4,57$; Тәжірибе жүзіндегі $\chi_t^2=4,57$ χ^2-тың кестедегі теориялық мәнінен үлкен, яғни $4,57>3,84$, сондықтан таңдама жиіліктің орта жиіліктен ауытқуы елеулі деп есептеледі және ол кездейсоқтық оқиғаға жатпайды.

2.3. Вариация коэффициенті
(құбылу коэффициенті)

Вариация коэффициенті де математикалық статистикада "орта квадраттық ауытқу" тәрізді жиіліктің құбылу жағдаятының өлшемі ретінде қолданылады.

Анықтама. Вариация (құбылу) коэффициенті деп "орта квадраттық ауытқу" шамасын "орта жиілікке" бөліп, ол бөліндіні *100* санына көбейткенге тең болатын шаманы атаймыз. Яғни бұл анықталған шама "орта квадраттық ауытқудың" "орта жиілікке" қатынасының пайыздық көрінісі деуге болады.

Егер "вариация коэффициентін" *"V"* деп белгілесек, онда:

$$V=\frac{\sigma}{\overline{X}}.100\%.$$

Бұл өрнектен ең алдымен байқайтынымыз, ол – орта квадраттық ауытқу шамасы неғұрлым көп және орта жиілік шамасы неғұрлым аз болса, вариация коэффициенті де солғұрлым үлкен шамаға ие болады.

Тәжірибеде, егер құбылу коэффициентінің шамасы 40 пайыздан көп болатын болса, таңдама жиіліктердің ауытқуы елеулі деп саналып, жиіліктің ауытқуының кездейсоқтығы жайлы жорамал теріске шығарылады. Әрине вариация коэффициенті *"V"* дәлдік жағынан "хи квадрат" критерийіне қарағанда төмендеу дәрежеде деп саналады.

Мысал ретінде жоғарыда сөз болған ғылыми-техникалық мәтініндегі етістіктердің жиіліктерін келтірейік: 98, 87, 102, 105, 123, 108, 85, 78, 110, 104; n=10; орта жиілік – $\overline{x}=100$.

Ал квадраттық ауытқу ортасы: $\sigma=\sqrt{\frac{\sum(x_i-\overline{x})^2}{n}}=12,7.$

Енді вариация немесе құбылу коэффициенті:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{12,7}{100} \cdot 100\% = 12,7\% .$$

Мұндай вариация коэффициенті жиіліктердің құбылу сипатының кездейсоқтығына айтарлықтай дәлел бола алады.

"Хи квадрат" критерийі мен "құбылу коэффициенті" мәндерінің ұтымды не ұтымсыз жақтарын тілдік және математикалық тұрғыда айқындай түсу үшін көптеген тәжірибелер жүргізу қажеттігін айта кеткенді жөн санаймыз.

2.4. Үлес ұғымы және оларды салыстыру

Анықтама. Үлес дегеніміз бақылауға қажетті "оқиғаның" барлық (осы тәріздес) қатарлар ауқымынан алатын орны (үлесі).

Үлес мәнін де ықтималдықты табу өрнегі арқылы анықтаймыз. Егер ықтималдықты латын әліпбиінің үлкен *"P"* әрпімен белгілесек, үлесті соның кіші әрпі *"p"* арқылы белгілеуге болады және ол ықтималдықтың классикалық анықтамасы бойынша (өрнегімен) есептеледі.

Есеп. *1000 сөзқолданыстан тұратын мәтін ішінде 300 етістік сөздердің барына көзіміз жетсе, онда етістіктің мәтін бойындағы үлесі қанша?*

$P=\frac{m}{n}=\frac{300}{1000}=0,3$ не болмаса пайыздық салмағын білу үшін оны *100-ге* көбейту керек, яғни $0,3\times 100\%=30\%$.

Жауабы: етістік сөздердің мәтін бойынан алатын үлесі *0,3* немесе ол барлық мәтіннің 30 пайызы.

Жиіліктердің ауытқуы тәріздес үлестер де құбылып тұрады. Үлестің ауытқуын табу үшін тағы да статистика пәніндегі "квадраттық үлес ауытқуы" деп аталатын өрнекті пайдалану қажет. Ол үшін үлес ауытқуы тек бір ғана статистикалық заңдылыққа бағынуы керек.

Егер *"M"* әрпімен квадраттық үлестің ауытқуын белгілесек, онда: $M=\sqrt{\frac{pq}{n}}$, бұл жердегі p – зерттелетін "оқиғаның" үлесі, q – барлық басқа оқиғалардың жиынтығының үлесі, сондықтан $q=1-p$, ал $p=1-q$ болады, n – таңдама жиынның көлемі.

Енді жоғарыда келтірілген есептің берілісі бойынша, мәтін бойындағы барлық етістік емес сөзқолданыстардың үлесін табуға болады. Яғни оны q деп белгілесек, онда $q=1-0,3=0,7$.

Квадраттық үлес ауытқуының өрнегі екі түрлі таңдама жиындардағы бірдей "оқиғаның" үлестерін салыстыру мақсатымен қолданылады.

Мысалы, Ахмет Байтұрсынұлының және Мағжан Жұмабаевтың шығармаларындағы "баяндауыш" сөздердің үлесін анықтағымыз керек болды делік не болмаса көркем әдебиет пен публицистика стильдеріндегі сын есім үлестерін білгіміз керек болды десек, осы "квадраттық үлес ауытқуы" өрнегіне аздаған өзгеріс енгізіп, пайдалануымызға болады.

Аталған өзгеріс мынада:

$$\varepsilon_{1,2} = \sqrt{\overline{pq}(\frac{1}{n_1}+\frac{1}{n_2})},$$

мұндағы $\varepsilon_{1,2}$ – квадраттық орта үлес ауытқу шамасының екі түрлі жиындар бойынша алынуы; $\bar{p}$ және $\bar{q}$ – анықтауға қажет және одан басқа да оқиғалардың жиынтығы бойынша есептелген орта үлестер; n_1 және n_2 – таңдама жиындардың көлемі.

Егер $3\ \varepsilon_{1,2} \le p_1\text{-}p_2$ болса, онда үлестердің ауытқуы елеулі дәрежеде болады, ал керісінше жағдайда – ауытқу кездейсоқ деп қорытынды жасау қажет.

Есеп: *1000 сөзқолданыстан тұратын екі түрлі мәтіндер жиындары бойынша етістіктердің кездесу жиілігі 200 және 150-ге тең.*

Етістіктің екі жиын бойынша анықталған үлестерінің статистикалық теңдігі жайлы ғылыми болжам жасауға болар ма еді?

Шешімі:

$$p_1=200:1000=0,20; \quad p_2=150:1000=0,15;$$

$$\bar{p}=\frac{p_1+p_2}{2}=(0,20+0,15):2=0,35:2=0,175;$$

$$q_1=1-0,20=0,80; \quad q_2=1-0,15=0,85;$$

$$\bar{q}=\frac{q_1+q_2}{2}=(0,80+0,85):2=1,65:2=0,825;$$

$$n_1=n_2=1000;$$

$$\varepsilon_{1,2}=\sqrt{0,175\cdot 0,825\cdot {}^{2}\!/_{1000}}=0,017.$$

Егер $3\times\varepsilon_{1,2} \le p_1\text{-}p_2$ болса, онда үлестердің ауытқуы елеулі дәрежеде болады, ал керісінше жағдайда – ауытқу кездейсоқ сипатта деп қорытынды жасау қажет.

Сонымен, қарастырып отырған жағдайда $3\times\varepsilon_{1,2}=0,051$ және $0,051\approx 0,050$. Бұл айырым $(0,501\text{-}0,050=0,001)$ аз болғанымен етістіктің екі жиын бойынша үлестерінің статистикалық теңдігі жайлы ғылыми болжам қабылданбайды. Яғни мұндай жағдайда үлестердің ауытқуы кездейсоқтық сипаттағы заңдылыққа бағынбайды, ауытқу елеулі деген қорытынды жасалады.

2.5. Орта таңдама жиіліктерді салыстыру

Таңдама жиіліктерді және олардың үлестерін салыстыруға болатын сияқты, "орта таңдама жиіліктер" де өзара салыстырылады.

Есеп: *500 сөзқолданыстан тұратын А және В мәтіндерінен он-оннан таңдама бөліктер алынған. Бұл мәтіндер жайлы іштей біркелкі деген үйғарым жасалды делік. Егер А және В мәтін таңдамаларында* ***сын есім сөздердің*** *жиіліктері төмендегідей болып кездессе:*

"А" бойынша: 72, 65, 78, 71, 70, 74, 80, 90, 68, 82;

"В" бойынша: 80, 93, 84, 83, 78, 67, 85, 86, 75, 89.

Онда осы "А" және "В" мәтіндері бойынша есептелген орта жиіліктер $\bar{X}_1$ *пен* $\bar{X}_2$ *араларындағы ауытқу шамасы кездейсоқтық сипатта ма, әлде елеулі дәрежеде ме?*

Бұл сұраққа жауап беру үшін математикалық статистикадағы "Стьюдент критерийін" пайдалану қажет [17, 29, 30, 40-43]. Оның өрнегі мынандай:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{1,2}} \cdot \sqrt{\frac{k_1.k_2}{k_1 + k_2}} \quad . \qquad (1)$$

Мұндағы $\bar{x}_1$, $\bar{x}_2$ – орта жиіліктер, k_1, k_2 – таңдама бөліктердің саны *(біздің мысалда $k_1=k_2=10$);* $S_{1,2}$ – квадраттық орта ауытқуларды бағалау параметрі. Аталған параметрді төмендегі (2) өрнекпен анықтайды:

$$S_{1,2} = \sqrt{\frac{\sum (x_{i1} - \bar{x}_1)^2 + \sum (x_{i2} - \bar{x}_2)^2}{k_1 + k_2 - 2}} \quad . \qquad (2)$$

Енді $\bar{x}_1$, $\bar{x}_2$, k_1, k_2, $S_{1,2}$ шамалары бойынша (1) өрнектегі t-ны тауып, сол мән арқылы ауытқудың кездейсоқтығы мен елеулілігінің ықтималдығын арнайы кесте (*2.2-кесте*) көмегімен анықтауға болады (осыған байланысты Б.Н.Головин "Язык и статистика" 40–50-бб. немесе басқа да оқулықтарды қараңыз).

Тақырыптың басында келтірілген есеп берілісіне қайта оралсақ, оның шешімі төмендегідей көрініс табады:

$$\bar{x}_1 = 75, \ \bar{x}_2 = 82, \ \Sigma (x_{i1} - \bar{x}_1)^2 = 508,$$

$\Sigma (x_{i2} - \bar{x}_2)^2 = 494$, $k = k_1 + k_2 - 2 = 10 + 10 - 2 = 18$; *$K=18$ – еркіндік дәреже саны;* K_1, K_2 – *тәжірибе саны.* Ал квадраттық орта ауытқуларды бағалау параметрі:

$$S_{1,2} = \sqrt{\frac{508 + 494}{18}} = 7{,}5.$$

Енді (1) Стьюдент критерийінің өрнегі бойынша:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{1,2}} \cdot \sqrt{\frac{k_1.k_2}{k_1 + k_2}} = \frac{82 - 75}{7{,}5} \cdot \sqrt{\frac{10 \cdot 10}{10 + 10}} = 0{,}93 \cdot \ 2{,}22 = 2{,}1.$$

2.2-кестедегі 18-жол бойынша $t=2{,}101$, ал бұл *5%* ықтималдыққа сай келеді. Бұл ықтималдық шамасы орта жиіліктерді ($\bar{x}_1$, $\bar{x}_2$), статистикалық тұрғыда тең екендігін терістеуге аз да емес, ал оны қолдауға айтарлықтай көп те емес. Яғни ауытқудың елеулі не кездейсоқ екендігін айту қиын, сондықтан қосымша байқау тәжірибесін жүргізген жөн болады.

Осы статистикалық есепті, яғни екі орта жиілікті салыстыруды басқа жолмен де шығаруға болады. Ол үшін квадраттық ауытқу мен олардың айырымын білу керек.

Аталған шаманы есептеп шығару үшін мына өрнек ыңғайлы деп саналады:

$$\varepsilon_{1,2} = \sqrt{\frac{\sigma_1^2}{\kappa_1} + \frac{\sigma_2^2}{\kappa_2}},$$

мұндағы σ_1^2 және σ_2^2 – екі таңдаманың дисперсиялары, K_1 , K_2 – тәжірибе саны немесе таңдама бөліктер саны.

Жоғарыда дисперсия мына өрнек арқылы берілгенін еске түсірейік:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{\kappa};$$

Есептеп шығарған квадраттық ауытқу мәні, яғни $\varepsilon_{1,2}$ екі орта жиіліктердің ($\overline{x}_1$, $\overline{x}_2$) айырымымен салыстырылады. Егер бұл айырым ($\overline{x}_1$ - $\overline{x}_2$) $\geq 3 \cdot \varepsilon_{1,2}$ болса, онда ауытқудың елеусіздігі жайлы болжам терістеледі.

Жоғарыда анықталғандар: $\overline{x}_1 = 75$, $\overline{x}_2 = 82$ және

$$\sigma_1^{\,2} = \frac{\sum (x_{x1} - \overline{x}_1)^2}{\kappa_1} = \frac{508}{10} = 50{,}8;$$

$$\sigma_2^{\,2} = \frac{\sum (x_{i2} - \overline{x}_2)}{\kappa_2} = \frac{494}{10} = 49{,}4.$$

Енді жаңадан ұсынылып отырған өрнек бойынша:

$$\varepsilon_{1,2} = \sqrt{\frac{50{,}8}{10} + \frac{49{,}4}{10}} = 3{,}17. \qquad 3 \cdot \varepsilon_{1,2} = 3 \cdot 3{,}17 = 9{,}51.$$

2.2-кесте

Стьюдент критерийіндегі *"t"* мәнін кесте арқылы анықтау

Еркіндік дәреже саны ($k=k_1+k_2-2$)	Үлкен шаманың ықтималдығы				
	0,50 (50%)	0,20 (20%)	0,10 (10%)	0,05 (5%)	0,025 (2,5%)
1	1,000	3,078	6,314	12,706	25,452
2	0,816	1,886	2,920	4,303	6,205
3	0,765	1,638	2,353	3,182	4,176
4	0,741	1,533	2,132	2,776	3,495
5	0,727	1,476	2,015	2,571	3,163
6	0,718	1,440	1,943	2,447	2,969
7	0,711	1,415	1,895	2,365	2,841
8	0,706	1,397	1,860	2,306	2,752
9	0,703	1,383	1,833	2,262	2,685
10	0,700	1,372	1,812	2,228	2,634
11	0,697	1,363	1,796	2,201	2,593
12	0, 695	1,355	1,782	2,179	2,560
13	0, 694	1,350	1,771	2,160	2,533
14	0, 692	1,345	1,761	2, 145	2,510
15	0, 691	1,341	1,753	2,131	2,460
16	0, 690	1,337	1,746	2,120	2,473
17	0, 689	1,333	1,740	2,110	2,458
18	0, 688	1,330	1,734	2,101	2,445
19	0, 688	1,328	1,729	2,093	2,433
20	0, 688	1,325	1,725	2,086	2,423
21	0, 687	1,323	1,721	2,080	2,414
22	0, 686	1,321	1,717	2,074	2,406
23	0, 685	1,319	1,714	2,069	2,398
24	0, 685	1,318	1,711	2,064	2,391
25	0, 684	1,316	1,708	2,060	2,385

Ал, $x_2 - x_1 = 82 - 75 = 7$, яғни 7 < 9,51. Яғни ол айырым $(X_2 - X_1)$ үш еселенген квадраттық ауытқудан $(3 \cdot \varepsilon_{1,2})$ кіші. Бұл жағдай екі орта жиіліктер ($\bar{x}_1$, $\bar{x}_2$) арасындағы ауытқудың елеулі еместігінің куәсі.

Сонымен, бірінші рет критерий Стьюдент бойынша күмән тудырған жайт, екінші әдіс бойынша анықталып, екі қатардағы жиіліктердің құбылуы бір ықтималдыққа бағынатын статистикалық заңдылық деп қорытындылауға болады.

2.6. Бақылау кезіндегі абсолюттік және қатынастық қателер мен таңдама мәтіннің көлемін анықтау

Статистикалық әдісті тілдік таңдама бөліктерге қолдану кезінде ықтималдық заңдылықтарға тән құбылмалықтың әсерінен зерттеушінің алған нәтижелері кейбір қателерге әкеледі. Сондықтан, мысалы, тілші анықтаған *орта таңдама жиіліктің* мәнін шындыққа сай келетін *"нағыз орта жиілік"* мәнімен және, сол сияқты, зерттеуші есептеп шығарған *"үлес"* мәнін *"нағыз үлес"* мәнімен салыстыруға бола ма? – деген сұрақ туындауы мүмкін.

Математикалық статистика пәні зерттеуші-тілшіге ондай жағдайды туғыза алады, яғни аталған шамалардың *"нағыз орта жиілік"* пен *"нағыз үлес"* мәндерінен құбылу шамасын дәл бір санмен көрсете алмағанымен, оның өзгеру шекарасын (ауқымын) анықтай алады. Ондай мүмкіндік мынандай *"бақылау қатесін"* анықтайтын өрнекке негізделеді [17, 29, 30, 40-50]:

$$L = \frac{t \cdot \sigma}{\sqrt{k}} \quad . \qquad (1)$$

Бұл өрнекті басқа атпен *абсолютті қате* деп те атайды. Мұндағы t – арнайы кесте арқылы анықталатын *теориялық коэффициент.* Оның мәні *"еркіндік дәреже санына"*, яғни, басқаша айтқанда, таңдама бөліктердің санына байланысты анықталатын өзгермелі шама. Ал, σ – *квадраттық ауытқу ортасы, k – таңдама бөліктер саны (бақылау саны).* Тәжірибе кезінде *квадраттық ауытқу ортасы* өрнегінің (σ-ның) орнына мынадай (2) өрнекті пайдаланған жөн деп саналады:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{k - 1}} \quad . \qquad (2)$$

Енді *"бақылау қатесін"* анықтау үшін қажетті шамалардың барлығы да бар. Олар: σ, k және t. Арнайы кесте арқылы анықталатын *теориялық коэффициент* t-ның мәні ықтималдықтың дәлдігіне байланысты алынады. Әдетте, үлестің орта жиілігін және байқау қатесін есептеуде 95% сенімділік жеткілікті деп есептеледі, яғни ол *0,95* ықтималдыққа сәйкес келеді деген сөз. Енді *"t"* шамасын *2.3-кестедегі* таңдама саны мен *95%-ға* сәйкес келетін жатық және тік жолдардың қиылысындағы санның шамасын теориялық коэффициент t-ның мәні деп есептейміз.

Мысалы, таңдама бөліктердің саны *к=5* болса, *t=2,78*, ал *k=10* болса, *t=2,26* тең деген сөз.

Кестеде көрсетілген коэффициент шамасы көп болған сайын, бақылау кезіндегі жіберген қате де дәлірек анықталады, яғни ықтималдық дәрежесі үлкен шама болады. Аса үлкен дәлдіктің қажеттігі болмаған жағдайда, тәжірибеде ондай коэффициентті кестесіз-ақ, *t=2* деп алуға да болады деп

есептейді. Ал, *95%* сенімділік дегеніміз жоғарыда көрсетілген өрнек бойынша есептелген бақылау кезіндегі қатенің шамасы *100 рет* жүргізілген статистикалық тәжірибеміздің бесеуінде ғана қайталануы мүмкін деген сөз. Осындай дәлдік мөлшері тілдерді зерттеуге байланысты жайттарда жеткілікті деп саналады.

Жоғарыда қарастырған есептің берілісін тағы да қайталайық.

Есеп: *А және В мәтіндерінен он-оннан таңдама бөліктер алынған. Бұл мәтіндер жайлы іштей біркелкі деген ұйғарым жасалды делік. Ал осы мәтін таңдамаларында* ***сын есім сөздердің*** *жиіліктері төмендегідей:*

"А" бойынша: 72, 65, 78, 71, 70, 74, 80, 90, 68, 82;

"В" бойынша: 80, 93, 84, 83, 78, 67, 85, 86, 75, 89.

Осы екі жиіліктер қатары бойынша анықталған орта таңдама жиіліктердің: $\bar{x}_1 = 75$, $\bar{x}_2 = 82$ *мәндерінің тәжірибе кезіндегі* "бақылау қатесін" *анықтаңыздар.*

Есептің шешімін табу үшін абсолютті қате өрнегіндегі ($L = \frac{t \cdot \sigma}{\sqrt{k}}$ - *t*), σ және *k* мәндері белгілі болуы керек.

Есептің берілісі бойынша бізге белгілі: $N=500$, $k=k_1=k_2=10$. *Квадраттық ауытқу ортасы* $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{k}}$ өрнегі арқылы анықталады және олардың шамалары:

$$\sigma_1 = \sqrt{\frac{508}{10}} = 7{,}1; \quad \sigma_2 = \sqrt{\frac{494}{10}} = 7.$$

2.3-кесте бойынша $k=10$ болса $t=2{,}26$ екені жоғарыда дәлелденген болатын. Олай болса:

$$L_1 = \frac{t \cdot \sigma_1}{\sqrt{k}} = \frac{2{,}26 \cdot 7{,}1}{\sqrt{10}} = \frac{16{,}046}{3{,}16} = 5{,}1;$$

$$L_2 = \frac{t \cdot \sigma_2}{\sqrt{k}} = \frac{2{,}26 \cdot 7}{\sqrt{10}} = \frac{15{,}82}{3{,}16} = 5{,}0.$$

2.3-кесте

"Бақылау қатесі" өрнегіндегі "t" мәнін кесте арқылы анықтау

Таңдама бөлік саны	Бақылау қатесін анықтаудың сенімділігі (ықтималдық)					
	99% (0,99)	97,5 (0,975%)	95% (0,95)	90% (0,90)	80% (0,80)	60% (0,60)
3	9,93	6,21	4,30	2,92	1,89	1,06
5	4,60	3,50	2,78	2,13	1,53	0,94
6	4,03	3,16	2,57	2,02	1,48	0,92
7	3,71	2,97	2,45	1,94	1,44	0,91
8	3,50	2,84	2,37	1,90	1,42	0,90
9	3,36	2,75	2,31	1,86	1,40	0,89
10	3,25	2,69	2,26	1,83	1,38	0,88
15	2,98	2,51	2,15	1,71	1,35	0,87
20	2,86	2,43	2,09	1,73	1,32	0,86
25	2,80	2,39	2,06	1,71	1,32	0,86
30	2,76	2,36	2,05	1,70	1,31	0,85

$L_1=5,1$ және $L_2=5$ – *бақылау қателері.* Сондықтан, бақылау қатесінің шамасын ескере отыра, орта таңдама жиіліктің мәнінің өзгеру шекарасын төмендегідей жазуға болады:

"А" мәтіні үшін: $75-5,1 \le \bar{x}_1 \le 75+5,1$ немесе $\bar{x}_1=75$ емес, оның мәні $69,9 \le \bar{x}_1 \le 80,1$ аралығында;

"В" мәтіні үшін: $82-5 \le \bar{x}_2 \le 82+5$ немесе $\bar{x}_2=82$ емес, оның мәні $77 \le \bar{x}_2 \le 87$ аралығында.

Ықтималдықтың *95%* сенімділікпен алынуы, орта жиіліктің есептелген аралықтан тыс жату мүмкіндігі *100* тәжірибенің бесеуінде ғана кездесуі мүмкін деген сөз.

Есеп: *А және Б мәтіндерінің көлемі N=500 сөзқолданысқа тең келетін және олар 5 таңдама бөлікке бөлінген екі түрлі мәтіндер бойынша* ***сын есім сөздердің*** *кездесу жиілігі төмендегідей:*

"А" – 55, 70, 76, 49, 45

"Б" – 52, 78, 88, 22, 25.

Тәжірибе кезіндегі бақылау қатесін табыңыз және "нағыз орта жиілік" қандай аралықта жататынын анықтаңыз.

Шешімі:

$$1)\ \bar{x}_1 = \frac{55+70+76+49+45}{5} = \frac{295}{5} = 59;$$

$$2)\ \bar{x}_2 = \frac{52+78+88+22+25}{5} = \frac{265}{5} = 53;$$

$$3)\ \sum_{i=1}^{5}(x_{i1}-\bar{x}_1)^2 = (55-59)^2+(70-59)^2+(76-59)^2+(49-59)^2+(45-59)^2 =$$

$$= 4^2+11^2+17^2+(10)^2+(-14)^2 = 16+121+289+100+196=722;$$

$$4)\ \sum_{i=1}^{5}(x_{i2}-\bar{x}_2)^2 = (52-53)^2+(78-53)^2+(88-53)^2+(22-53)^2+(25-53)^2 =$$

$$= 1+625+1225+961+784 = 3596.$$

5) Енді *ауытқудың квадраттық ортасын* анықтайық:

$$\delta_1 = \sqrt{\frac{\sum(x_{i1}-\bar{x}_1)^2}{K_1}} = \sqrt{\frac{722}{5}} = 12,00,$$

$$\delta_2 = \sqrt{\frac{\sum(x_{i2}-\bar{x}_2)^2}{K_2}} = \sqrt{\frac{3596}{5}} = 27.$$

6) *95% сенімділік* үшін, $k=5$ болса, *2.3-кесте* бойынша $t=2,78;$

7) *бақылау қателері*:

$$L_1 = \frac{t\cdot\delta_1}{\sqrt{k}} = \frac{2,78\cdot 12}{\sqrt{5}} = \frac{33,36}{2,24} = 14,9;$$

$$L_2 = \frac{t\cdot\delta_2}{\sqrt{k}} = \frac{2,78\cdot 12}{\sqrt{5}} = 33,7;$$

8) *"А"* мәтіні үшін $\bar{x}_1$ ауытқуларының аралығы:

$59-14,9 \le \bar{x}_1 \le 59+14,9;$ $\quad 44,1 \le \bar{x}_1 \le 73,9;$

9) "Б" мәтіні үшін: $19,3 \leq \bar{x}_2 \leq 86,7$;

"Б" мәтініне байланысты ауытқу аралығы өте үлкен, сондықтан мұндай "жағымсыз" жағдай болмас үшін не мәтін көлемін, не ішкі таңдама бөліктер көлемін ұлғайту, не *"t"* мәнін кішірейту қажет болады.

Жоғарыда келтірген (1) өрнек, яғни $L=\frac{t \cdot \sigma}{\sqrt{k}}$ өрнегі бақылаудың *абсолютті қатесін* анықтау үшін қолданылады, себебі оның өлшемі таңдама орта жиіліктің өлшемімен бірдей болып келеді.

Бақылаудың *абсолютті қатесін* анықтау барлық жағдайларда зерттеушіні қанағаттандыра бермейді. Сол себепті статистиктер өз тәжірибесінде *абсолютті қателеден* басқаша – *"қатынастық қате"* деп аталатын шаманы да қолданады. Егер *абсолютті қате* таңдама орта жиіліктің *"нағыз орта жиіліктен"* не көп, не аз болып ауытқуын білдіретін санға тең болатын болса, *қатынастық қате* – осы *абсолютті қатенің таңдама орта жиілікке* қатынасына тең болатын ондық бөлшек (немесе пайыздық шама).

Мысалы, абсолютті қате $L=25$, ал орта жиілік $\bar{x}=50$ десек, онда *қатынастық қатені* "δ" *(дельта)* деп белгілесек:

$$\delta=\frac{L}{\bar{x}}=\frac{t \cdot \sigma}{\bar{x} \cdot \sqrt{k}}=\frac{25}{50}=\frac{1}{2}=0,5.$$

Яғни *"абсолютті қате" "орта жиіліктің"* $\frac{1}{2}$ бөлігін немесе *50* пайызын құрайды екен.

Ал $L=25$ және $\bar{x}=500$, $\delta=\frac{25}{100}=\frac{1}{20}=0,05$, яғни абсолютті қате "$L$" орта жиіліктің $\frac{1}{20}$ бөлігін немесе *5* пайызын құрайды екен.

Қатынастық қатені есептеу арқылы *абсолютті қате* мен *орта жиілікті* бір-бірімен салыстыруға мүмкіндік туады, яғни қате шамасының қай жағдайда – көп, ал қай жағдайда – аз екендігін ажырата аламыз.

Статистикалық жолмен тіл заңдылығын зерттеу тәжірибесінде қатынастық қатенің *5-10%* шамаларын қанағаттандырарлық деп санайды. Кейбір жағдайларда *30%-ға* дейінгі шамамен де келісуге болады, ал одан көбі көптік етеді. Сонымен, *қатынастық қатені* есептеу өрнегі:

$$\delta=\frac{t \cdot \sigma}{\bar{x} \cdot \sqrt{k}} \quad . \qquad (3)$$

Бұл өрнектегі $\frac{t \cdot \sigma}{\sqrt{k}}=L$ – *абсолютті қате*, $\bar{x}$ – *орта жиілік.*

Егер зерттеуші "оқиғаның" *орта жиілігін* емес, оның барлық зерттеу мәтінінен алатын *үлесін* білгісі келсе, ол төменде көрсетілетін өрнектер арқылы іске асады. Яғни тәжірибе жүзінде есептелген *"үлестің" "нағыз үлестен"* ауытқуын білу үшін *абсолютті* және *қатынастық қателерді* анықтау қажет болады.

Ал ондай қателер (4) және (5) өрнектер арқылы есептеледі:

$$L_p=\frac{2\sqrt{pq}}{\sqrt{N}}. \qquad (4)$$

Бұл (4) өрнек бойынша үлестің *абсолютті қатесін* табуға болады. Мұндағы "*2*" – тұрақты коэффициент, *p* және *q* – таңдама үлестері, *N* – сөз санымен не басқа тілдік бірлік арқылы өлшенетін мәтін көлемі.

Келесі (5) математикалық өрнек арқылы *қатынастық қатені* анықтауға болады:

$$\delta_p = \frac{2\sqrt{q}}{\sqrt{pN}}. \qquad (5)$$

Қатынастық қатені анықтайтын (5) өрнек статистикалық тәжірибе үшін таңдама бөліктердің санын және олардың көлемін алдын ала анықтауға мүмкіндік туғызады.

Мәтін көлемін анықтау үшін мынадай амалдарды іске асырайық:

1) (5) өрнектегі теңдіктің екі жағын квадратқа алсақ:

$$\delta_p{}^2 = \frac{4q}{pN}, \qquad (6)$$

2) енді осы (6) өрнек бойынша мәтін көлемін анықтайтын (7) формуланы жазуға болады:

$$N = \frac{4q}{p\delta_p^2}. \qquad (7)$$

Бұл өрнектердегі *p* – зерттеуге тиісті оқиғаның барлық мәтін бойынан алатын үлесі, ал *q* – басқа оқиғалардың үлесі, δ_p – осы үлеске қатысты *абсолютті қате* шамасы. Егер осы параметрлердің шамаларын алдын ала өз қалауымызша белгілесек, оларға сай мәтін көлемін (7) өрнек бойынша есептеп шығаруға болады.

Осындай мәтін көлемін *абсолютті қате* өрнегі бойынша да табуға болады:

$$L = \frac{t \cdot \sigma}{\sqrt{k}}; \quad L^2 = \frac{t^2\sigma^2}{k}; \quad k = \frac{t^2\sigma^2}{L^2}.$$

Егер таңдама бөлік саны *k=N болса,* яғни ол барлық мәтін көлеміне тең десек, онда бұл өрнек мәтін көлемін есептеп шығаруға қолданылады.

Төмендегі өрнектерді де мәтін көлемін анықтау үшін қолдануға болады:

а) *абсолютті қате* (*L*) бойынша мәтін көлемін анықтау:

$$N = \frac{4p \cdot q}{L^2};$$

б) *қатынастық қате* (δ) бойынша мәтін көлемін анықтау:

$$N = \frac{4 \cdot q}{\delta^2 \cdot p}.$$

Есеп: *Алдыңғы есептің берілісі бойынша мәтін ішіндегі үстеу сөздердің үлесі 0,07. Егер абсолютті қате 0,005-тен аспау үшін қандай көлемді мәтін алғанымыз жөн.*

Шешімі:

p=0,07; q=1-0,07=0,93; L=0,005;

$$N = \frac{4 \cdot 0{,}93 \cdot 0{,}07}{(0{,}005)^2} = 10{,}416;$$

Немесе мұны $N=10500$ сөзқолданыс деп шамамен (дөңгелетіп) жазуға болады.

Есеп: *1-ші есеп берілісі бойынша үстеу үлесі 0,07 болудың және қатынастық қатенің $\delta=0,05$ аспау үшін мәтін көлемі "N" қандай болу керек*

Шешімі:

$p=0,07;\ q=0,93;\ \delta=0,05;$

$N=\frac{4\cdot q}{б^2\cdot p}=\frac{4\cdot 0,93\cdot 0,07}{(0,005)^2}=21,257$ сөзқолданыс, яғни шамамен алғанда $N=21000$ сөзқолданыс болуы керек деп саналады.

Үшінші тарау
ҚАЗАҚ МӘТІНІН ЫҚТИМАЛДЫ-СТАТИСТИКАЛЫҚ МОДЕЛЬДЕУ

3.1. Ципф заңы және оны қазақ мәтіні бойынша түзілген жиілік сөздіктерге қолдану

Ципф заңының қысқаша тарихы. 1916 жылы француз ғалымы Эсту стенографиялық жазу ісін жетілдіру мақсатындағы тәжірибе үстінде мынадай заңдылықты байқаған болатын [37].

Мәселен, көлемі *N*-ге тең сөзқолданыс құрайтын мәтін бойынша түзілген жиілік сөздіктің реестрлік сөздер (сөзтұлғалар) саны *n*-ге тең болып, ол *3.1-кестедегідей* көрініс тапты делік.

3.1-кесте

Рет саны *(i)*	Реестрлік сөз (сөзтұлға) *(S_i)*	Абсолютті жиілік *(F_i)*	Қатынастық жиілік
1	S_1	F_1	P_1
2	S_2	F_2	P_2
----	---	---	----
i	S_1	F_i	P_i

Егер осы сөздіктегі кез келген реестрлік сөздің рет санын (нөмірін) "*i*", ал оған сәйкес келетін сөздің абсолютті жиілігін F_i деп белгілесек:

$1 \le i \le n$; $F_i = F_1, F_2, F_3, \ldots, F_n$ және $F_i \ge F_{i+1}$, $P_i = \frac{F_i}{N}$ – қатынастық жиілік болатыны анық.

Сонда француз ғалымы Эсту мынадай жағдайдың куәсі болады: $F_i \cdot 1 \approx F_2 \cdot 2 \approx F_3 \cdot 3 \approx \ldots \approx F_n \cdot n$, яғни $F_i \cdot i \approx F_{(i+1)} \cdot (i+1)$.

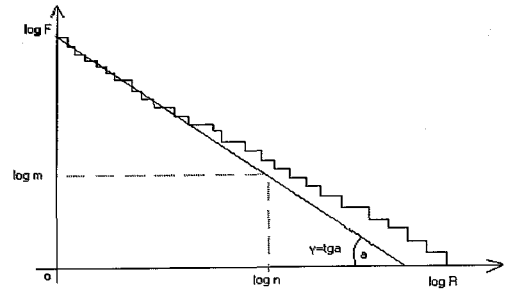
Бұл өрнектің мағынасы мынада: жиілік сөздіктегі кез келген сөздің (сөзтұлғаның) абсолютті жиілігі мен оның рет санының көбейтіндісі, шамамен алғанда, бір-біріне тең келеді. Егер ондай тең шамаларды тұрақты "*C*" санымен белгілесек:

$$F_i \cdot 1 \approx F_2 \cdot 2 \approx F_3 \cdot 3 \approx \cdot \cdot \cdot \approx F_n \cdot n = C, \qquad (1)$$

яғни $F_i \cdot i \approx F_{(i+1)} \cdot (i+1) = C$, $F_i = \frac{C}{i}$, $(i=1, 2, 3, 4, 5, 6, \ldots, n)$.

Сонымен, (1) өрнек мына заңдылықты аңғартады: жиілігі кему тәртібімен орналасқан жиілік сөздікте сөздің рет саны ұлғайған сайын, оның кездесу жиілігі керісінше кеміп отырады, яғни сөздің жиілігі мен оның рет саны бір-бірімен өзара байланыста және кері пропорционалдық заңдылықта болады.

Ал 1928 жылы телефон компаниясының қызметкері Э.Кондон француз ғалымы Эстуге қатыссыз, мына заңдылықты ашады: егер билогарифмдік координат жүйесіндегі абсцисс өсіне жиілік сөздіктегі сөздердің рет санының *(i)* ондық логарифмдік шамасы $(\ell g i)$, ал ординат өсіне сөздердің абсолют жиілігінің ондық логарифмдік шамасы $(\ell g F_i)$ салынатын болса, онда өстер жазықтығындағы нүктелер жүйесі – сынық сызықтардан тұратын сызба бейнесі төмендегі *1-суретте* көрсетілгендей түзу сызықтық пішінге ұқсас болып шыққан.



1-сурет

$F_i=\frac{c}{i}$ өрнегі билогарифмдік координатта мына түрде көрініс табады:

$$F_i=\frac{c}{i^{\gamma}}. \qquad (2)$$

Бұл (2) өрнектегі C және γ – тұрақты сандар (константтар), ал $\gamma=tg\alpha$ – түзу сызық пен абсцисс өсінің арасындағы α бұрышының тангенсіне тең, C – түзу сызықтың ордината өсімен қиылысатын нүктесінің ординат шамасына тең.

Енді (2) теңдіктің екі жағын да бірдей N санына бөлетін болсақ: $\frac{F_i}{N}=\frac{c}{N\cdot i^{\gamma}}$, мұндағы $\frac{F_i}{N}=P_i$ – қатынастық жиілік, ал $\frac{C}{N}=K$ – тұрақты сан, яғни $P_i=\frac{k}{i^{\gamma}}=k\cdot i^{-\gamma}$. (3)

Осы (3) өрнек – *Ципф заңы* деп аталады. Бұл заңдылық бойынша, мәтіннің кез келген жерінен таңдалып алынған сөздің сөздіктегі орны, яғни рет саны зерттеушінің қалаған нөміріне тең болу ықтималдығы анықталады.

Ципф заңының мағыналық жағы мынада. Белгілі мәтін мен оның жиілік сөздігі бойынша K мен γ мәндерін бір-ақ рет анықтап алып, одан кейін кез келген сөздің сөздіктегі рет саны арқылы ол сөздің қолдану жиілігін анықтау мүмкіндігі және керісінше, сөздің абсолютті жиілігі бойынша ол сөздің сөздіктегі рет санын анықтау. Абсолютті жиілік мәні жоғарыда көрсетілген (3) өрнек бойынша былайша анықталады:

$$F_i=N\cdot k\cdot i^{-\gamma}. \qquad (4)$$

Соңғы (4) көріністегі өрнек те Ципф заңы деп аталып жүр.

Ақпарат теориясы мен математикалық әдістердің XX ғасырдың 50-ші жылдары тілтаным ғылымына қарқындап араласуы көпшілік ғалымдардың Ципф заңына қызығушылығын арттырды. Әсіресе, Ципф заңының бірнеше тілдердегі көрінісі мен тұрақты коэффициенттер шамасының әр тілге қатысты құбылуының мағыналық түсінігіне және т.б. осыған байланысты жағдайларға ғалымдар назар аудара бастады.

Ципф заңындағы (4) γ константасының мәтіннен мәтінге ауысқан сайын тұрақтылық сипатынан айырылатынын белгілі ғалым Б.О.Мандельброт өзінше дәлелдеп, ол заңды былай жазуды ұсынады [65]:

$$F_i = N \cdot K\,(i+\rho)^{-\gamma}. \qquad (5)$$

Мұндағы ρ – тұрақты шама.

Енді осы соңғы жазылған (5) өрнек – Эсту-Ципф-Мандельброт заңдылығы деп атала бастады.

Р.М.Фрумкина орыс тілі мәтіндері негізінде γ шамасының тәжірибе үшін алынған бір мәтін ішінде де тұрақты бола бермей, өзгеріске ұшырайтынын дәлелдеді [90]. Ғалымның осы заңдылыққа қатысты тұжырымдаулары төмендегідей:

1. Эсту-Ципф-Мандельброт (5) өрнегіндегі γ – константасы тек белгі шекаралықта (интервалда) ғана тұрақты болады, ал одан сырт жағдайда γ рет саны *i*-ге тәуелді және өспелі функция. Мысалы, рет санының *50* мен *1500* аралығында, яғни $50 \le i \le 1500$ аралығында γ мәні тұрақты, ал рет санының одан сырт мәндерінде γ өзгеріске ұшырайды және оның мәні *"i"* шамасына қатысты өсіп отырады.

2. Әр мәтінге байланысты γ-ның тұрақтылық сақтайтын рет саны бойынша алынған шекарасы өзгеріске ұшырайды және ол мәтіннің статистикалық құрылымына тікелей қатысты.

3. γ шамасы мәтін көлемінің өзгеруіне қарай да құбылып отырады, яғни $\gamma = f(N)$. Сондықтан екі (не одан да көп) мәтіндерді салыстырып зерттеу кезінде ондай мәтіндердің көлемдері тең (не шамалас) болуы шарт.

Тәжірибе жүзінде Эсту-Ципф-Мандельброт заңдылығындағы γ, K және ρ мәндерін анықтау үшін мына теңдеулер жүйесін пайдалануға болады:

$$\begin{cases} F_1 = N \cdot K\,(i_1+\rho)^{-\gamma}, \\ F_s = N \cdot K\,(i_s+\rho)^{-\gamma}, \\ F_{соңы} = N \cdot K\,(i_{соңы}+\rho)^{-\gamma}. \end{cases} \qquad (6)$$

Мұндағы F_s сөздік үзіндісінің екі шеткі абсолютті жиіліктері – F_1 мен $F_{соңы}$ шамаларының геометриялық ортасы, ал i_s – рет сандары i_1 мен $i_{соңы}$ шамаларының геометриялық ортасы. K мен γ мәндерін алдын ала белгілі ρ бойынша "ең кіші квадраттар әдісі" деп аталатын әдіспен анықтауға болады.

Теңдеулер жүйесіндегі параметрлердің мәнін оңай табу үшін қажетті деген компьютерлік бағдарлама жасалуы қажет. Қазақстандық статолингвистер кезінде өз есеп-санақтары үшін А.В.Зубов пен Э.Н.Хотяшовтың ЭЕМ-ға арнап құрастырған алгоритмдерін пайдаланған болатын [47].

М.Әуезовтің "Абай жолы" романы көлемдері шамалас 4 томнан (кітаптан) тұратыны және роман мәтінінің көлемі 465966 сөзқолданысты құрайтыны жайлы екінші тарауда айтылған болатын. Әрбір кітаптың мәтіндерін бөлек-бөлек қарастырып, олардан түзілген жиілік сөздіктер бойынша есептеліп шығарылған Эсту-Ципф-Мандельброт заңдылығы параметрлерінің салыстырма деректері *3.2-кестеде* берілді.

3.2-кесте

Эсту-Ципф-Мандельброт заңдылығы параметрлерінің салыстырма мәндері

Ципф заңының параметрлері	1-кітап	2-кітап	3-кітап	4-кітап	Роман "А.Ж."
Мәтін көлемі (N_i)	105788	124398	112727	123053	465966
$i_{сөн.}$ ($F_i \geq 4$)	3270	3872	3570	4027	9614
ρ	4,770	5,490	4,760	6,380	7,835
γ	0,942	0,937	0,924	0,930	1,017
К	0,092	0,088	0,079	0,086	0,146

Кесте деректері жоғарыда Р.М.Фрумкинаның γ мәнінің мәтін көлеміне байланысты өзгеріске ұшырайтынын дәлелдей түседі. Көптеген зерттеушілердің ұйғаруынша, γ параметрінің шамасы, сол мәтіннің не "шағын тіл" (подъязык) мәтінінің лексикалық әртүрлілік көрсеткіші ретінде түсіндіріледі [90].

Мысалы, γ шамасы үлкен мәнге ие болған сайын, сол тілдің лексикалық қорының кедейлігі не керісінше, γ мәнінің "кіші" болуы лексикалық қордың байлығы деп саналады. "Абай жолы" романының әр томына қатысты есептелген ($\gamma<1$) γ бірден кіші, ал балалар әдебиеті тілі бойынша жасалған жиілік сөздік деректерінде ол мән керісінше, яғни онда γ бір санынан үлкен шама қабылдайды екен (γ=1,6).

Жалпы алғанда, әр тіл немесе әр стиль бойынша есептелген γ мәндері не тіл типінің, не стильдер типінің айырымдық белгілері болады деп тұжырымдауға негіз бола бермейді. Яғни бұл параметр әмбебаптық қасиетке ие емес, ол тек лексикалық бірліктің жиілік сөздіктегі рет санына қарай өзгеретін шама деуге болады. Осыған көз жеткізу үшін "Абай жолы" романы жиілік сөздіктерінің өн бойында γ мәні қалайша өзгеріске ұшырайтынына қысқаша тоқталайық.

Төменде *3.3-кестеде* көрсетілгендей, романның әрбір кітабының жиілік сөздіктерінің рет саны 10 түрлі пішінде екі-екі бөліктерге бөлінген. Яғни бірінші ондықта бастапқы бөлу нүктесінде $i=1$, ал екінші ондықта соңғы нүктеде $i=i_{соңы}$.

Әрбір бөлу нәтижелері бойынша Ципф заңының γ параметрі есептеліп, олардың шамалары айтылған кестеде көрініс тапты. Тіке сызықтық регрессия коэффициенттері арқылы әрбір кестеде көрсетілген зоналар бойынша және арнайы жазылған компьютерлік бағдарлама көмегімен γ мен $lg(N{\cdot}K)$ мәндері есептелді (келесі тақырыпшада берілген *3.3* пен *3.4-кестелерді* қараңыз).

Кестеде көрініс тапқан γ мәндері басқа тілдер бойынша жасалған тұжырымдарды дәлелдей түседі. Яғни γ мәні сөздіктің әр бөлігінде әртүрлі және олардың сандық мәні *1* санына шамалас келеді. γ мәтіннің кейбір *1* санынан барынша алшақтау мәндері сөздіктің бастапқы және соңғы бөліктеріне сәйкес келеді, ал ортаңғы бөлігінде γ параметрінің тұрақтылық қалпы байқалады.

"Абай жолы" романының толық мәтіні бойынша γ мәтінінің ортаңғы бөліктерінің арифметикалық орта шамасы *0,913* тең. Сол сияқты Ципф заңы өрнегіндегі K мәнін де тұрақты деуге бола бермейді, себебі сөздіктің ортаңғы бөлігімен салыстырғанда, шеткі бөліктерінде біршама ауытқулар байқалады.

Сонымен, қорыта айтқанда, Ципф заңдылығы үндіеуропа тілдерінен түзілген жиілік сөздіктері бойынша қалай орындалса, негізінен алғанда, түркі тілі, соның ішінде қазақ тілі мәтіндерінде де сол дәрежеде орындалады деуге әбден болады.

Егер $F_i=N{\cdot}K(i+\rho)$ заңдылығындағы K, ρ, γ параметрлері белгілі болса, онда әрбір реттік санға *(i)* қатысты абсолютті жиілік $(F_i^{эм})$ үшін, оған сәйкес келетін теориялық абсолютті жиілікті (F_i^T) анықтауға болады. Ал бұл Ципф заңының "тік сызықтық" пішіндегі теориялық жиіліктер үлестірілуін тәжірибе жүзінде алынған (эмпиризмдік) жиіліктер үлестірілуі арасындағы алшақтықты бағалауға мүмкіндік тудырады.

Әрбір рет саны i-ге сәйкес келетін $F_i^{эм}$ мен F_i^T мәндерін салыстыру орта квадраттық ауытқу өрнегі арқылы жүзеге асады:

$$\sigma = \sqrt{\frac{1}{q} \cdot \sum_i \left(F_i^{эм} - F_i^T\right)^2} \quad . \qquad (7)$$

Мұндағы q – байқау саны. Мысалы, "Абай жолы" романы мәтіндері бойынша түзілген жиілік сөздіктер бойынша орта квадраттық ауытқу мәндері мынадай:

$$\sigma_1=18,280;\ \sigma_2=15,737;\ \sigma_3=15,550;\ \sigma_4=17,192$$

және толық роман мәтіні бойынша $\sigma_5= 66,261.$

3.2. Жиілік сөздіктегі ранг пен жиілік арасындағы корреляция

Тақырып атындағы "ранг" терминін жиілік сөздіктегі сөздердің рет саны деп, ал "корреляция" терминін екі өзгермелі шама арасындағы өзара байланыстылық не арақатыстылық деп түсінген жөн.

Екі өзгермелі шама, яғни жиілік сөздіктегі сөздің (сөзтұлғаның) рангі мен жиілігі арасындағы байланыстың тығыздылығын немесе "күшін" және түрпатын (түрін) корреляциялық талдау негізінде анықтауға болады.

Егерде аталған байланыс түрін "түзу сызықтық" деп ұйғарсақ, онда оның өрнегі $y=ax+b$ түрінде көрініс табу керек. Теңдеудегі $y=lgF_i$, $x=lgi$ билогарифмдік өстегі сызықтық функцияның ординат және абсцисс координаттары, i – ранг, F_i – рангке сәйкес сөздің абсолютті жиілігі. Яғни теңдеудің билогарифмдік өсіне қатысты түрі: $lgF_i=a\cdot lgi+b$ және оның жазықтықтағы сызбасы алдыңғы тақырыпшадағы *(1-суреттегі)* көрінісімен бірдей келеді.

Ендігі мақсат – тік сызықтың $y=ax+b$ теңдеуіндегі a және b параметрлерін анықтау. Қойылатын шарт – эмпиризмдік және теориялық сызықтар арасындағы кез келген екі жұп нүктелердің квадраттанған арақашықтықтары ең кіші дәрежеде болуы қажет.

Алға қойған мақсатқа жету үшін "ең кіші квадраттар әдісін" қолданып, ранг және жиілік арасындағы тәжірибе жүзіндегі – сынық сызық түрпаттағы кескінін "түзу сызықтық регрессия теориясы" бойынша сызықтық функция түріне келтіру керек. Ал регрессия теориясы ол үшін a мен b мәндерінің ең қолайлы деген мәндерін табуға мүмкіндік туғызады [2].

Мұндай жағдайда регрессия түзуінің теңдеуі былай жазылады:

$$y_x-\bar{y}=R\cdot\frac{\sigma_y}{\sigma_x}\cdot\left(x-\bar{x}\right), \qquad (1)$$

мұндағы $y_x=lgF_i$; $x=lgi$, x пен y билогарифмдік өске қатысты lgi мен lgF_i мәндерінің арифметикалық ортасы, ал σ_x пен σ_y – орта квадраттық ауытқулар, R – x пен y арасындағы байланыстың күшін анықтайтын корреляция коэффициенті.

Енді (1) теңдеуді $y=ax+b$ түріне келтіру үшін y_x анықтайтын болсақ, ол төмендегі (2) пішінді қабылдайды:

$$y_x=R\cdot\frac{\sigma_y}{\sigma_x}\cdot x+\left(\bar{y}-\bar{x}\cdot R\cdot\frac{\sigma_y}{\sigma_x}\right), \qquad (2)$$

ал a мен b мәндері: $a=R\cdot\frac{\sigma_y}{\sigma_x}$ және $b=\bar{y}-\bar{x}\cdot R\cdot\frac{\sigma_y}{\sigma_x}$.

3.3-кесте

"Абай жолы" романы жиілік сөздіктері бойынша Ципф заңындағы γ параметрі мен *R* корреляция коэффициентінің мәндері

Бөліктер	1-кітап		2-кітап		3-кітап		4-кітап	
	γ	*R*	γ	*R*	γ	*R*	γ	*R*
1-15	0, 97	0,97	0,47	0,94	0,48	0,97	0,38	0,99
1-50	0,98	0,98	0,59	0,98	0,59	0,99	0,56	0,98
1-100	0,99	0,99	0,62	0,99	0,63	0,99	0,59	0,99
1-300	0,99	0,99	0,71	0,99	0,70	0,99	0,69	0,99
1-1000	0,82	0,99	0,99	0,99	0,81	0,99	0,79	0,99
1-1500	0,85	0,99	0,84	0,99	0,84	0,99	0,82	0,99
1-2000	0,88	0,99	0,86	0,99	0,86	0,99	0,84	0,99
1-3000	0,91	0,99	0,89	0,99	0,88	0,99	0,87	0,99
1-4000	0,93	0,99	0,91	0,99	0,91	0,99	0,89	0,99
1-соңы	0,92	0,97	0,93	0,99	0,89	0,97	0,93	0,98
15-соңы	0,93	0,97	0,93	0,98	0,89	0,97	0,93	0,97
50-соңы	0,93	0,97	0,93	0,97	0,89	0,97	0,93	0,97
100-соңы	0,92	0,97	0,93	0,97	0,89	0,97	0,93	0,97
300-соңы	0,92	0,97	0,93	0,97	0,89	0,96	0,93	0,97
1000-соңы	0,89	0,95	0,91	0,96	0,86	0,95	0,91	0,96
1500-соңы	0,86	0,94	0,89	0,95	0,84	0,94	0,87	0,95
2000-соңы	0,83	0,93	0,87	0,94	0,81	0,93	0,89	0,94
3000-соңы	0,77	0,90	0,82	0,92	0,76	0,91	0,82	0,92
4000-соңы	0,67	0,86	0,76	0,89	0,70	0,87	0,77	0,89

Арнайы жазылған компьютерлік бағдарлама бойынша *x*, *y*, σ_x, σ_y, *R* шамаларын есептеп шығаруға болады. Енді *a* және *b* мәндері арқылы Ципф заңындағы γ және *K* параметрлерін де анықтауға болады.

3.4-кесте

"Абай жолы" романы толық мәтіні бойынша Ципф заңының γ параметрі, *lg(N·K)* және *R* корреляция коэффициенті

Зонаның р/с	Шекаралық бөліктер	γ	*lg(N·K)*	*R*
1	1 –15	0,457	3,824	0,961
2	1 – 50	0,599	5,801	0,983
3	1 – 100	0,633	9,183	0,992
4	1 – 1000	0,815	4,271	0,995
5	1 – 4000	0,914	4,514	0,997
6	1 – 6000	0,950	5,616	0,996
7	1 – 10000	1,001	4,770	0,996
8	1 – 15000	1,045	4,912	0,995
9	1 – 25000	1,090	5,066	0,994
10	1 - соңы	1,065	4,951	0,982
11	15 - соңы	1,068	4,964	0,982
12	50 - соңы	1,071	4,978	0,981
13	100 - соңы	1,073	4,988	0,981
14	1000 - соңы	1,079	5,014	0,975
15	4000 - соңы	1,045	4,860	0,958
16	6000 - соңы	1,011	4,704	0,945
17	100000 - соңы	0,925	4,311	0,911
18	15000 - соңы	0,782	3,650	0,842
19	25000 - соңы	0,271	1,272	0,447

Ол үшін, $F_i=N\cdot K\cdot i^{\gamma}$ теңдеуінің екі жағы бірдей логарифмделсе жеткілікті:

$$lgF_i= -\gamma\cdot lgi+lg\,(N\cdot K).$$

Егер $y=lgF_i$ және $x=lgi$ екенін ескерсек, $a= -\gamma$ және $b=lg(N\cdot K)$ болатыны анықталады да, ал соңғы екі теңдіктен γ мен K мәндерін оңай табуға болады.

Жиілік сөздіктердегі сөздерді өз қалағанымызша зоналарға (бөліктерге) бөлуге болады және әр зонаға сай γ мен K мәндерін анықтауға мүмкіндік бар.

Алдыңғы тақырыпта қарастырған "Абай жолы" романы бойынша түзілген әр кітаптың және толық роман мәтінінің жиілік сөздіктерінің бөліктеріне қатысты есептелген корреляция коэффициенттері екі түрлі кестелер *(3.4* және *3.5)* арқылы көрініс тапты.

3.3-кесте мен *3.4-кестеде* келтірілген корреляция коэффициенттері жайлы деректерді қарастыра келе, мынадай қорытынды жасауға болады: түзу регрессия кезінде корреляция коэффициентінің *(R)* мәні *y* пен *x* шамаларының түзу сызықтық байланысының тығыздығын білдіретін және оның "күшін" бағалайтын параметр ретінде қарастырылады.

Сонымен, егер:

а) $0 \le R \le 1$ болса, *x* пен *y* аралығындағы байланыс – "тікелей байланыс" (прямая зависимость) болып саналады;

ә) $-1 \le R \le 0$ болса – "кері байланыс" (обратная зависимость) болып саналады;

б) корреляция коэффициентінің *(R)* мәні *1* санына жуықтаған сайын аталған байланыстың тығыздығы күшейе түседі.

Кестеде көрсетілген деректер бойынша (*R*-дің мәндері), "Абай жолы" романының әр кітабының жиілік сөздіктеріндегі сөзтұлға жиілігі мен оның рангісінің (реттік санының) арасындағы байланыс – кері байланыс екендігін $(-1 \le R \le 0)$ және олардың айтарлықтай "күшті" еместігін атауға болады. Корреляция коэффициентінің ең үлкен *(-0,994÷-0,996)* мәндері жиілік сөздіктің орта тұстарына сәйкес келетіндігі анықталды. Сонымен бірге, сөздіктердің бас жағына қатысты байланыс күші, оның соңғы жақтарымен салыстырғанда күштілеу (тығыздау) болатындығы да байқалады. "Абай жолы" романы мәтінінен түзілген сөздіктердегі сөзтұлғалардың рет саны (ранг) мен жиіліктері арасындағы корреляциялық қатынастылық күшінің тығыздығын корреляция коэффициентінің арифметикалық ортасын анықтап та білуге болады:

$$\overline{R}=\frac{\sum_{j=1}^{19}R_j}{19}. \qquad (3)$$

"Абай жолы" романының әрбір томы мен олардың қосындысы (роман) бойынша есептелген корреляция коэффициентінің орта шамасы мен жиіліктің тәжірибелік және теориялық мәндерінің орта квадраттық ауытқулары төмендегідей *(3.5-кестеде)* көрініс тапты:

3.5-кесте

Корреляция коэффициентінің орта шамасы, тәжірибелік және теориялық жиілік мәндерінің орта квадраттық ауытқулары

"Абай жолы" романы мәтіндері	1-том	2-том	3-том	4-том	Роман
$\overline{R}$	- 0,966	-0,969	-0,967	-0,971	- 0,943
σ_R	0,00033	0,00028	0,00033	0,00028	0,00014

Корреляциялық коэффициент мәнінің сенімділігіне баға беру (оценка надежности) үшін "квадраттанған қатенің ортасын" анықтайтын мына өрнекті (формуланы) қолдануға болады [20, 35]:

$$\sigma_R = \frac{1 - R^2}{\sqrt{N}}. \qquad (4)$$

Мұндағы R – жиілік сөздік бойын түгел қамтыған жағдайдағы корреляция коэффициенті, N – жиілік сөздік бойындағы әртүрлі сөзтұлғалар (сөздер) саны. Егер N айтарлықтай көп болса, онда корреляция коэффициентінің ақиқаттық шекаралық мәні мынадай интервал аралығында жататындығы дәлелденді:

$$R - 3\sigma_R \le R_{ақиқ} \le R + 3\sigma_R.$$

Біз қарастырып отырған "Абай жолы" романының сөздіктерінде мұндай интервалдар төмендегідей:

$-0{,}975 \le R_1 \le -0{,}974;$
$-0{,}977 \le R_2 \le -0{,}975;$
$-0{,}973 \le R_3 \le -0{,}972;$
$-0{,}976 \le R_4 \le -0{,}975;$
$-0{,}982 \le R_{ром.} \le -0{,}981.$

Сонымен, корреляция коэффициентін сенімділік дәрежеге бағалау нәтижелері R мәнінің жоғары дәрежедегі дәлдік болатындығына дәлел. Ал роман материалдары бойынша алынған сөздіктердің өн бойына тән қасиет – корреляциялық байланыстың түзу сызықты сипатта болуы және ондай байланыстың сөздік бойының орта жағында "күшті", ал шеткі жақтарында "әлсіз" болатындығы.

3.3. Мәтін және оның жиілік сөздігі ішіндегі сөздің (сөзтұлғаның) ақпараттық сипаттамасы

Тіл – қатынас құралы. Сондықтан оның элементтеріне ақпараттық өлшем жүргізілуі қажет және оған тиісті баға берілуі керек деп ұйғарылады.

Осындай пайымдаулардан мынадай тұжырымдар туындайтыны анық:

1. Тіл арқылы қатынас жасау (речевое общение) – тілді байланыс торабына (каналына) жатқызуды қажет етеді. Ақпараттық байланыс тілдің әріп, дыбыс, морфема және басқа тілдік бірліктері арқылы іске асады.

2. Тіл арқылы қатынас жасау үшін тілдік бірліктерді белгілі бір "кодтың" символдары ретінде қарастыру керек. "Код" дегеніміз – "тіл", бірақ шектеулі тіл. Себебі, бұл жағдайда тілдегі тіркесімдер мен басқа да бірліктердің қолдану ықтималдығына қажетті шектеулер қойылады.

3. Тілдік қатынас кезінде хабарламаны тарату мен оны қабылдау аралығын қосу – сөйлеу каналы (торабы) арқылы іске асады. Бұл жағдайда айтушы мен қабылдаушы бірдей "кодты" пайдаланулары қажет.

Синхрондық лингвистика саласындағы әрбір зерттеушінің көздейтін мақсаты хабарлама ішіндегі лингвистикалық бірліктердің көрінісін реттейтін тілдік "кодтағы" шектеулерді анықтау және бағалау. Осындай бағалаудың ең бір ұтымды деп саналатын әсерлі әдісі, ол – лингвистикалық бірліктер мен қатынастарды осы тәріздес ақпараттық өлшем – энтропия, яғни артықтық (избыточность) сияқты сандық дәрежедегі теориялық-ақпараттық шамалармен салыстыру.

Тілді ақпараттық тұрғыда бағалау мүмкіндігі тек нақты тілде жазылған мәтінді статистикалық зерттеу арқылы ғана іске асады. Мәтін арқылы таратылатын ақпарат мөлшерін анықтауда ондай мәтін белгілі бір "кодқа" енетін дискретті бірліктер тізбегі ретінде қарастырылады. Дискретті тілдік бірлік – ол сөйлемдер, сөздер, буындар, фонемалар, әріптер. Осы аталған бірліктердің әрбіреуінің мәтін бойынан орын алуының көрінісі тәжірибе тізбегінің қайталану ретімен іске асады. Бұл жердегі "тәжірибе жүргізу" дегеніміз – бірліктер жиыны ішінен осы мәтін бөлігінде шығуы мүмкін болатын қайсыбір мәтін бірлігін таңдап алу деп түсінген жөн. Осылайша таңдалып алынған әрбір бірлік – айнымалы X шамасының әртүрлі ықтималдықпен қабылданатын мәндерінің бірі деп саналады.

Белгісіздік сипаттағы осындай айнымалы X-тің сандық мәні "энтропия" терминімен аталады. Айнымалы X-тің нақты мәніне сілтеме жасау дегеніміз, ол энтропияны (белгісіздікті) жою және сондай мөлшердегі ақпарат алу. Бұл түсінікте ақпараттың маңыздылық (семантикалық) жағы сөз болуы мүмкін емес. Сондықтан мұндай жағдайда тек селективті ақпарат жайлы сөз болады. Яғни мәтін бөлігіндегі ақпараттың лингвистикалық бірліктерін таңдау ықтималдық жүйе арқылы іске асатындығы ескерілуі қажет. Іс жүзінде тілдің синтаксис пен лексика салаларында элементарлы бірліктер саны (сөйлемдер, сөзтіркестер, сөздер) шексіздікке дейін өсуі мүмкін. Сол сияқты, тілдің фонетика саласындағы фонемаларды (дыбысты) дискретті бірліктерге ажырату проблемасының да шешілмеген жақтары баршылық. Сондықтан да теориялық-ақпараттық зерттеулерде, көбінесе, әріптік код көрінісіндегі және шекті дискретті бірліктен тұратын жазба тіл материалдары пайдаланылады. Осыған байланысты тілдегі ақпараттық өлшеуді іске асыру – мәтін бойындағы әр әріпке қатысты энтропияны анықтаудың әрекеті. Әрбір мағыналы мәтін өзара қатынастық сипаттағы және әр дәрежелі ықтималдықпен анықталатын әріптер тізбегі екені мәлім. Бірақ энтропияның әрбір әріпке шаққанда келетін "шындық мәніне" (истинное значение – H) жуық келетін мәнін табу үшін мәтін әріптері бірдей ықтималдықта және өзара тәуелсіз сипатта болу қажет деген ұйғарым жасалуы керек.

Әрине, энтропиялық өлшемді іске асыру мәселесін толық түрде қарастыру бұл жұмыстың мақсатына жатпайды. Сондықтан бұл мәселе жайлы жайттарды толығырақ білгісі келген оқырман "Энтропия языка и статистика речи" (Минск, 1966) [99] атты

жинақпен және "Әдебиет" тізімінде көрсетілген басқа да ғылыми еңбектермен танысуына болады [76].

Келесі қарастырмақшы мәселеміз М.Әуезовтің "Абай жолы" романы мәтінінен түзілген "сөзтұлғалар жиілік сөздіктері" бойынша қазақ көркем әдебиет мәтініндегі бір сөзтұлғаға түсетін орташа ақпарат мөлшерін анықтау. Ол үшін бірдей дәрежедегі бірлікке түсетін "орташа ақпарат мөлшерінің" орташа энтропияға *(H)* тең болатындығын ескере отырып, К.Шеннонның белгілі формауласын пайдалануға болады:

$$I = H = -\sum_{i=L}^{i_{кон}} f_i \cdot \ell og_2 f_i \,, \qquad (1)$$

мұндағы f_i – қатынастық жиілік i – сөзтұлғаның сөздіктегі рет саны, $f_i \cdot \ell og_2 f_i$ – көбейтіндісі арқылы сөздіктегі кезекті рет саны үшін хабар бірлігіндегі "өлшенген ақпарат" (взвешенная информация) мөлшері анықталады. Ал ақпарат мөлшері *(I=H)* екілік санау жүйесіндегі "бит" өлшемімен анықталады.

Аталған (1) өрнекті қолдану үшін, жиілік сөздіктегі сөзтұлғаларды мәтін бойынан рет-ретімен жорамалмен таңдалып алынатын тәжірибе нәтижелерінің жиынтығы деп қарастыру қажет. Әрбір тәжірибе нәтижесі: мәтін ішіндегі кездесетін сөзтұлғаның қатынастық жиілігіне тең болатын нақты ықтималдық шамасын анықтау.

"Өлшенген жиынтық ақпарат" шамасын есептеу үшін (1) өрнекті басқаша түрде (жиілік сөздіктің *1* санынан *n* саны аралығын қоса есептейтіндей) қайта жазсақ:

$$I^* = H^* = \sum_{i=1}^{n} f_i \cdot \ell og_2 f_i \,. \qquad (2)$$

Бұл өрнек бойынша, сөзтұлға жиілігі мен "өлшенген ақпарат" санының тікелей қатыстығын анықтай аламыз. Ауызша не жазбаша мәтіндегі сөзтұлғаның қолдану жиілігі ұлғайған сайын, оның "өлшенген ақпарат" саны да артады. Басқаша айтқанда, сөзтұлғаның мәтінге "енгізетін" ақпараты арта түскен сайын, оның қолданысы да көбейе түседі және мәтін ішіндегі басқа сөздермен тіркесімдігі (немесе мәнмәтіндік қоршауы) де жиіліктік сипатта болады (және керісінше).

Сол сияқты, (2) өрнек арқылы жиілік сөздіктің әр бөлігіне қатысты "жиынтық өлшенген ақпаратты" және оның барлық сөздік бойынан алатын пайыздық үлесін де анықтауға мүмкіндік туады. Бұл жердегі пайыздық үлес – "ақпараттық салмақ" (информационный вес) деген терминмен аталып, мына түрде жазылады:

$$\eta = \frac{I^*}{I} \cdot 100\ \% \,. \qquad (3)$$

Мұндағы I – бір сөзтұлғаға қатысты "орташа ақпарат саны", ал I^* – бір топ сөзтұлғаларға қатысты жиынтық ақпарат салмағы.

М. Әуезовтің "Абай жолы" романы мәтінінің жиілік сөздігі бойынша бір топ сөзтұлғаның мәтінді қамту пайызына сәйкес жиынтық ақпарат саны мен ақпараттық салмақтың өсу қарқыны *3.6-кестеде* көрініс тапты. Бұл кестеде романның толық мәтіні мен 4-ші кітабы мәтінінің жиілік сөздіктеріндегі сөзтұлғаларға қатысты I^* мен η мәндері сөзтұлғалардың мәтін бойын әртүрлі *(1*-ден *100*-ге дейінгі*)* аралықтағы пайыздық қамту шамаларына сәйкестендіріліп берілген. Пайыздық қамту қарқыны артқан сайын, жиынтық ақпарат салмағы *(I*)* шамасы да сондай

шапшаңдықпен өседі. Мысалы, сөзтұлғаның 100% қамтуына *4*-ші кітап бойынша *12,5* дв.ед. ал роман бойынша – *12,7* дв.ед. сәйкес келеді. Бұл мәндер, орта шамамен алғандағы, бір сөзтұлғаға сәйкес келетін ақпарат мөлшерін анықтайды.

Жиілік сөздіктің алдыңғы жағынан соңына қарай есептеу нәтижесінде сөзтұлғалардың мәтінді қамту қарқыны ақпараттық салмақтың η өсу қарқынымен салыстырғанда шапшаңдау екенін байқауға болады. Сирек қолданыстағы сөзтұлға топтарына сәйкес келетін мәтінді қамту пайызы *(Q)* мен ақпараттық салмақ η мәндері және олардың қосынды шамалары *3.7-кестеде* берілді. Осы кестедегі мәліметтер бойынша, егер жиі қолданыстағы сөздер үшін "ақпараттық салмақ" пайызы η "мәтінді қамту" пайызынан *(Q)* кіші болатын болса, сирек қолданатын сөздер үшін, керісінше, үлкендік сипатта.

Екі кестеде де әртүрлі көлемдегі мәтіндерден *(V_4=123053; $V_{ром.}$=465966)* түзілген жиілік сөздіктер мәліметтері *(I^*, η)* қарастырылды. Оларды салыстыра қарастырудан шығатын қорытынды – сөздік бірлігінің ақпараттық өлшем дәрежесіне мәтін көлемінің айтарлықтай әсері болмауы.

3.6-кесте

М.Әуезовтің "Абай жолы" романы мәтінінің жиілік сөздігі бойынша бір топ сөзтұлғаның мәтінді қамту пайызына сәйкес жиынтық ақпарат саны мен ақпараттық салмақтың өсу қарқыны

Мәтінді қамту пайызы	4-кітап мәтіні бойынша		Роман мәтіні бойынша	
	I^*	η	I^*	η
1	0,0715	0,5727	0,0769	0,6036
5	2,2871	2,2996	0,3168	2,4867
10	0,7475	5,9874	0,7316	5,7427
15	1,1704	9,3748	1,1404	8,9515
20	1,6311	13,0650	1,5968	12,5183
25	2,1323	17,0796	2,0797	16,3246
30	2,6515	21,2383	2,6017	20,4220
35	3,2318	25,8865	3,1375	24,6277
40	3,7798	30,2759	3,7144	29,1561
50	4,9863	39,9399	4,9153	38,5825
60	6,2581	50,1270	6,2170	48,8041
70	7,7264	61,8879	7,6722	60,2227
80	9,2713	74,2625	9,2539	72,6383
100	12,4845	100,0	12,7397	100,0

Сөзтұлғаларды жеке және топтап қарастырып, олардың функционалдық мәнін анықтауда ақпараттық салмақ – тұрпайы сипаттағы баға деуге болады. Себебі, мұнда сөйлеу қызметі кезіндегі сөз бен сөз тіркестері арқылы байланысқа қатысты селективті ақпараттың 20–30 пайызын құрайтын лексика-грамматикалық ақпарат шамасы ескерілмей отыр. Солай бола тұра, көптеген зерттеушілердің пайымдауынша, сөздердің мәтінді қамту пайызымен салыстырғанда, "ақпараттық салмақ" мәні лексикалық бірліктердің функционалдық қарым-қатынасын мағыналы түрде сипаттайды.

3.7-кесте

Жиілік F_i	4-кітап мәтіні бойынша				Роман мәтіні бойынша			
	Q	$\sum Q$	η	$\sum \eta$	Q	$\sum Q$	η	$\sum \eta$
1	2	3	4	5	6	7	8	9
1	13,56	13,56	17,59	17,59	7,48	7,48	11,06	11,06
2	6,33	19,89	7,72	25,31	3,75	11,23	5,25	16,31
3	4,37	24,26	5,13	30,44	2,74	13,97	3,71	20,02
4	3,41	27,67	3,90	34,34	2,18	16,97	2,88	22,90
5	2,80	30,47	3,14	37,48	1,94	18,09	2,52	25,42
6	2,43	32,90	2,67	40,15	1,52	19,61	1,95	27,37
7	1,96	34,86	2,12	42,27	1,41	21,02	1,77	29,14
8	1,93	36,79	2,06	44,33	1,35	22,37	1,67	30,81
9	1,61	38,40	1,70	46,03	1,23	23,60	1,52	32,33
10	1,39	39,79	1,45	47,48	1,00	24,60	1,30	33,63

3.4. Лингвистикалық болжамды тексеру (сынау) критерийі

Қолданбалы лингвистиканың көптеген маңызды мәселелерінің бірі – мәтін ішіндегі тілдік бірліктердің орналасу үлгісінің (моделінің) ықтималды-статистикалық әмбебап сызбасын құру. Мұндай күрделі істі автоматты түрге келтіру: ақпаратты іздеу, тілдерді тиімді жолмен үйрету, базалық (негізгі) тілді құру, тілдің "нормалық" табиғатын түсіну, тілді стилистика тұрғысынан зерттеу және т.б. кешенді мәселелерді шешуде аса қажетті деп саналады [92].

Тілді математикалық модельдеу кезінде оның негізгі мәселесі болып саналатыны: тәжірибе жүзінде алынған сандық деректерге статистикалық талдау жұмысын жүргізу. Аталған талдау арқылы кейбір дәстүрлі зерттеу кезінде тікелей байқауға жатпайтын бірқатар тілдік жайттардың ара-жігі анықталады. Ал тілдің ықтималдық үлгісі көмегімен мәтін ішіндегі қайсыбір тілдік бірліктердің сандық қасиеттерін сипаттауға мүмкіндік туады. Егер осы анықталған бірліктер сипаттамасы ақиқат шындықтың көрсеткіші ретінде қабылданса, онда модельден нақты нысанға өту іске асады. Ол үшін зерттеуге алынған мәтін бірліктерінің сандық параметрлерін анықтай білу қажет. Бұл параметрлерге "арифметикалық орта" $(\bar{X})$ мен "стандартты" (S) жатқызуға болады. Әрі қарай $\bar{X}$ пен S мәндері "ықтималдық модельдің" осы тәріздес параметрлерімен салыстырылады.

Тілдік бірліктердің сипаттамалары зерттеушіні сол тәжірибеге түскен шекті көлемдегі нақты мәтінге ғана қатысты қызықтырмайды, ол бас мәтін жиынтығы сипаттамасы тұрғысынан қарастырғанды қалайды. Бұл жердегі "бас мәтін жиынтығы" (генеральная совокупность текстов) дегеніміз тілдің толық түрі немесе бөлек бір стиль, тіл қабаттары, жазушы тілі және т.б. Бұлайша қарастырудың себебі аталған "бас мәтін жиынтығын" қамтитын толық түрдегі мәтін бойынша статистикалық талдау жүргізу тәжірибе жүзінде мүмкін бола бермейді. Яғни зерттеуге жататын нысан "бас мәтін жиынтығының" таңдалып алынған жеке бөліктері (частные выборки) немесе "таңдама бөліктер". Міне, осындай таңдама мәтін бөліктер параметрлерін бас мәтін жиынына қатысты статистика-математикалық

зерттеу негізінде ықтималды модельден тілдік нысанның өзіне көшу (өту) жүзеге асады. Мұндай "өту" процесі нақты (белгілі) критерийлерге сүйеніп, статистика-лингвистикалық болжамды сынау негізінде бірліктердің (параметрлердің) мәтін ішінде кездейсоқ оқиға ретінде "үлестірілу" (таралу) тұрпатын (түрін) анықтау арқылы іске асады.

Мәселен, мынадай статистика-лингвистикалық болжамды сынау (тексеру) қажет болды делік: мәтін ішіндегі зат есімге қатысты сөздер математикалық статистиканың "нормальды үлестірілу" деп аталатын теориялық заңдылығына бағынады. Мұндай болжамды сынау (тексеру) үшін зат есім сөздердің эмпирикалық (тәжірибелік) үлестірілу деректерінің теориялық үлестірілуі "нормальды" заңдылығымен сәйкестік дәрежесін тексеру қажет. Ал ондай тексеру (сынау) тек қана зат есім сөздердің таңдама бөліктер ішінде қайталану жиілігінің орта шамасының бірнеше "сериялары" (мүмкіндіктері) бойынша қарастырылады.

Кездейсоқтық жағдайдағы лингвистикалық шаманың мәтін бойында таралу (үлестірілу) тұрпатының не параметрлерінің сынауға ұсынылып отырған болжамға сәйкестігі нөлдік (негізгі) болжам (H_0) немесе оған қарама-қарсы қойылатын альтернативті болжам (H_1) ретінде тұжырымдалуы мүмкін. Негізгі – нөлдік (H_0) болжам белгілі критерий арқылы сыналады. Критерийлер – статистикалық және реттік деп екіге ажыратылады. Алғашқысы (статистикалық) кездейсоқ лингвистикалық шаманың үлестірілуіне (қолдану жиілігіне) бағыттала отырып, нөлдік болжамды сынауға қолданылады, ал екіншісі аталған шамалардың реттік үлестірілуін қарастырады.

Әдетте, лингвистикалық болжамдар кездейсоқ таңдама бөліктер негізінде тексеріледі. Бірақ бұл тілдік бірліктің бас жиындағы үлестірілу көрінісінің дұрыстығының айғағы бола алмайды, себебі, мұндай жағдай "болжамды сынау" процесін жалған жолға бұрып жіберуі де мүмкін. Сондықтан дұрыс шешім орнына, мысалы, дұрыс нөлдік (H_0) болжам орнына, бұрыс (H_1) болжам (не керісінше) қабылданады. Осындай жайттарға байланысты зерттеушінің қай болжамды $(H_0$ не $H_1)$ негізгі деп қабылдауының да көп мәні бар.

Осы айтылғаннан жіберілетін қателердің мәнділік дәрежесі екі түрлі типпен ажыратылады. Олар – 1-ші және 2-ші дәрежелік мәнді қателер деп аталады. Тіл білімінде қатенің мәнділік дәрежелерін таңдау, көбінде, субъективті (жеке бастық) ой қорыту негізінде іске асады. Мысалы, түркі, моңғол, тунгус-маньчжур тілдерінің "алтайлық гипотеза" деп аталатын генетикалық туыстығы жайлы болжам белгілі статистикалық критерий арқылы тексеруге болатындай тұжырымдалған [54].

Мұндағы нөлдік болжам бойынша барлық алтай тілдері бір ғана генетикалық туыстас жеріне, яғни негіз-тілге (язык-основа) келіп тіреледі екен (восходят). Осы тұжырымды теріске шығаратын өзіне нағыз кәміл сенетін алтайшы да аса қауіпті 1-ші дәрежелік мәндегі қате жіберген болар еді. Бұл жерде нөлдікке қарсы қабылданған болжамның өзі дұрыс болмаса да, мәнділік жағынан 2-ші дәрежелі қате қауіпсіздеу деуге болады.

Сондықтан, болжамды сынау критерийін таңдау, сонымен бірге сынау аясын таңдаумен ұштасады деуге болады.

Сөйтіп қандай да болсын дұрыс шешім қабылдар тұсында 2-ші типті қате жіберуден гөрі 1-ші типті қате жібермеу аса маңызды деп саналады. Қате дәрежесінің мәнділігін азайту мақсатымен болжамдар өзара сыналуы қажет.

Тәжірибелік құндылығына қарай, яғни 1-ші типті қатеден күтілетін салдарына қарай, зерттеушіні қанағаттандыратын маңыздылық деңгейі таңдалады. Басқаша

айтсақ, жіберген қатеден туындайтын салдардың сипатына қарай маңыздылық деңгейін де азайтқан (төмендеткен) жөн болады.

Әдетте, лингвистикалық зерттеулерде маңыздылық деңгейінің $\alpha = 0,05$ немесе $\alpha = 0,01$ шамасы да жеткілікті деп саналады. Мәселен, алтай тілдеріне қатысты болжамды статистикалық тәсілмен тексеру (сынау) үшін көрсетілген маңыздылық деңгей шамасы қанағаттандырарлық деп саналады.

3.5. Қазақ мәтінінің ықтималды-статистикалық үлгісін (моделін) құру

Лингвистикалық статистика қазір өзінің алғашқы кезеңін, яғни лингвистикалық құбылыстарды сипаттауды аяқтап, келесі күрделі кезеңіне өтуде. Бұл – тілдік заңдылықтарды сандық тұрғыда алдын ала болжай алатын теория құру мәселесі.

Сонымен, лингвистикалық бірліктердің мәтін ішінде үлестірілуінің математикалық тұрпатын анықтау мәселесі тілдік құбылыстың моделін құрудағы ең бір маңызды іске айналып отыр [93].

Аталған мәселе бойынша қазақ тілі мәтіні ішінде сөз таптарының статистикалық үлестірілуінің кейбір теориялық заңдылықтарға сәйкес келу болжамын сынауға қатысты зерттеулер жүргізілді. Теориялық үлестірілу заңдылықтары ретінде математикалық статистикадағы "нормальды үлестірілуі", "Пуассон үлестірілуі" және "Шарлье үлестірілуі" заңдылықтары қарастырылды.

Қазақ тіліндегі мәтін нысаны ретінде М.Әуезовтің "Абай жолы" романынан 10 мың сөзқолданыс, газет мәтінінен 50 мың сөзқолданыс таңдама бөліктер тәжірибе ретінде алынды. Ал сөз таптары ішінен ең негізгілері деп: *зат есім, етістік, сын есім* және *үстеу* сөздердің мәтін ішіндегі статистикалық үлестірілуі сынаққа түсті.

Белгілі бір лингвистикалық бірліктің (мысалы, белгілі бір сөз табының) тілдегі қолданылу сипатын танып білу кезінде құбылу заңдылығы бұзылмайтын ең кіші (минимум) мәтін көлемін анықтау зерттеу ісінің маңызды шартына айналып отыр.

Осындай мақсатпен жүргізілген тәжірибе реті төмендегідей:

1) мәтін ішіндегі сөздерді сөз табына ажырату және оларды шартты белгілермен сәйкестендіру;

2) мәтіннің жалпы көлемі 25, 50, 100, 200 және 500 сөзқолданыстан тұратын микробөліктерге, яғни сериялық таңдама бөліктерге бөлу;

3) өзгермелі шамаларды шартты белгілеуді ұстану: әр стиль бойынша алынған мәтіннің жалпы көлемі – N; микробөлік (серия) көлемі – k; әр серияға сай микробөлік саны – n;

4) әр жағдайдың жалпылама жазылу пішінін қабылдау – $B_N^n = k$ (теңдік белгісі шартты). Мұндағы $N = n.\ k$. Мысалы, Абай жолы" романы мәтіні бойынша алынған 10000 сөзқолданыстан тұратын мәтін көлемін үшке бөліп: $N_1 = 2500$, $N_2 = 5000$, $N_3 = 10000$ қарастырсақ, әрбір N_i көлемді $k = 25, 50, 100, 200, 500$ сөзқолданыстан тұратын микробөліктерге (n) бөлсек, олардың саны өзгеріп отырады. Мәселен, егер $k = 25$ болса:

$N = 2500$ болса, $B_{2500}^{25} = 100$ бөлікке тең;

$N = 5000$ болса, $B_{5000}^{25} = 200$ бөлікке тең;

$N = 10000$ болса, $B_{10000}^{25} = 400$ бөлікке тең.

Сол сияқты, егер $k=50$ болса: $B^{50}_{2500}=50$; $B^{50}_{5000}=100$; $B^{50}_{10000}=200$ немесе егер $k=100$ болса: $B^{100}_{2500}=25$; $B^{100}_{5000}=50$; $B^{100}_{10000}=100$.

Газет мәтіні бойынша тәжірибеге алынған *50000* сөзқолданыстан тұратын мәтін көлемі де, жоғарыда көрсеткендей, серия бөліктеріне бөлініп зерттелді;

5) әрбір вариантқа қатысты микробөліктер ішіндегі қажетті сөз табының жиілігі саналып, бірдей жиілікпен кездесетін сөз табының серия саны анықталды. Осы деректердің негізінде жиілік пен микробөлік санынан тұратын төмендегідей дискретті вариациялық үлестірілімдік қатар түзілді:

x_i	x_1	x_2	x_3	...	x_s
n_i	n_1	n_2	n_3	...	n_s

Мұндағы x_i *(i=1,2,3,...,s)* сөз табының жиілігі, n_i *(i=1,2,3,...,s)* – әрбір x_i-ге сай микробөлік саны, $\sum_{i=1}^{s} n_i = n$.

Жүргізілген тәжірибе бойынша, көркем әдебиет пен газет мәтіндеріндегі негізгі сөз таптарының мәтін бойында таралуының вариациялық қатарларын (статистикалық үлестірілудің) кейбір теориялық заңдылықтармен үйлесімділік деңгейі зерттелді. Теориялық заңдылықтар ретінде Пуассонның, Шарльенің *(А, В түрлері)* және нормальдық, логарифмды-нормальдық үлестірілу заңдылықтары алынды.

Дискретті және үздіксіз вариациялық қатар құрудың мысалы ретінде "Абай жолы" романы мәтінінің *N=10000 сөзқолданыстан* тұратын таңдама бөлігін қарастырайық. Егер ішкі бөліктер (микровыборки) немесе серия таңдамаларының ұзындығы *k=100* сөзқолданыс десек, серия саны $B^{n}_{N}=k=B^{100}_{10000}=100$ болады екен. Енді әр сериялық мәтін ішіндегі *"зат есім"* сөздердің жиілігін анықтап, олардың ең аз қолдану саны – *19,* ал ең көп қолдану саны – *43* екенін білдік делік. Әрі қарай, *19* бен *43* сандары аралығын немесе *(19; 43)* интервалы аралығын *"1"* санына (жиілікке) өсіру негізінде мынадай дискретті вариациялық қатар құруға болады:

x_i	*19*	*20*	*21*	*22*	*23*	*24*	*25*	...	*40*	*41*	*42*	*43*	n_i қосындысы:
n_i	*1*	*0*	*2*	*3*	*3*	*3*	*5*	...	*1*	*0*	*0*	*1*	*100*

Жоғарыда аталған теориялық үлестірілу заңдылықтардың *Пуассон* және *Шарлье-В түрі* дискретті вариациялық қатар құру арқылы, ал *нормальды, логарифмды-нормальды* және *Шарлье-А түрі* үздіксіз вариациялық қатар құру негізінде олардың теориялық жиіліктері мен статистикалық параметрлері анықталды.

Дискретті вариациялық қатардан үздіксіз вариация қатарына көшу үшін Δx-ке тең шама (қадам) таңдалып алынып, $(x_i;\ x_{i+\Delta x})$ интервалдық қатар құрылады. Интервалдың шекаралық мәндерінің арифметикалық ортасы – x_j, және олардың әрқайсысына сай келетін n_j, табуға болады:

$$x_j v = \frac{x_i + x_{i+\Delta x}}{2}, \qquad (1)$$

$$n_j = \sum_{k=i}^{i+x_{i+\Delta x-1}} n_k \,. \qquad (2)$$

Мұндағы $i=1,2,3, \ldots, n$ – дискретті қатардың рет саны, $j=1,2,3,\ldots, m$ – үздіксіз қатардың рет саны.

Егер $\Delta x=3$, болса үздіксіз вариациялық қатар төмендегі *3.8-кестеде* көрсетілгендей болады:

3.8-кесте

$(x_i; x_{i+}\Delta_x)$	*(19; 22)*	*(22; 25)*	*(25; 28)*	...	*(40; 43)*	*(43; 46)*
x_jv	*20,5*	*23,5*	*26,5*	...	*41,5*	*44,5*
n_jv	*3*	*9*	*12*	...	*2*	*1*

Сонымен, осылайша құрған дискретті және үздіксіз вариациялық қатарлар мәтінге қатысты сөз таптары жиіліктерінің эмпирикалық үлестірілу заңдылығын анықтайды. Бұл деректер арнайы компьютерлік бағдарламалар негізінде теориялық жиілік пен статистикалық параметрлерді есептеп шығару үшін қажет.

Енді жоғарыда аталған теориялық үлестірілу заңдылықтардың екі ғана түрінің өрнектеріне (формуласына) тоқталайық.

3.5.1. Пуассонның үлестірілу заңдылығы

Лингвистикалық зерттеулерге қатысты ең бір маңызды теориялық заңдылық – Пуассонның үлестірілу заңдылығы. Бұл заңдылықты, көбінде, тәжірибе кезінде күтілетін әр оқиғаның шығу (пайда болу) ықтималдығы *(p)* өте аз, ал тәжірибенің саны барынша көп болған жағдайда қолданады.

Осы қойылатын шарттар қазақ тілінің мәтіндерінде кездесетін сөз таптарына қатысты да орындалады деп жорамал жасауға болады.

Пуассон заңдылығын қолдану үшін, жоғарыда айтылғандай, сөз таптарының статистикалық деректерінің негізінде дискретті вариациялық қатар құруымыз керек болады. Тәжірибе кезінде қалаған оқиғаның шығуын, мысалы, *зат есім сөздердің* кездесу мүмкіндігін (шамасын) *x*, "математикалық күтуді" (математическое ожидание) – *a* немесе оның орнына жүретін арифметикалық ортаны – $\bar{X}$, тәжірибе санын – *n* деп белгілесек, теориялық жиілік санын анықтайтын Пуассон заңының өрнегі төмендегідей жазылады [85]:

$$n^T = \frac{a^x \cdot e^{-a}}{xi} \cdot n. \qquad (3)$$

Бұл (3) өрнектегі "математикалық күтудің" орнына жүретін арифметикалық ортаның $(\bar{X})$ өрнегі мына түрде жазылады:

$$a \approx \bar{x} = \frac{1}{n} \cdot \sum n_i \cdot x_i' \,. \qquad (4)$$

3.5.2. Нормальды үлестірілу заңдылығы

Нормальды үлестірілу заңдылығы статистикалық зерттеулерде кеңінен қолданады деуге болады. Оның негізгі себебі, бұл заңдылықтың сызба пішінінің аса қарапайымдылығы және ол заңдылықтың көптеген *"жалғыз төбелік"* сипаттағы үлестірілулерге пішіндік жағынан жақындығы. *Нормальды үлестірілу заңдылығының* тағы бір ұтымды жері – тәжірибелік байқау кезіндегі деректерге қатысты әртүрлі математика-статистикалық өңдеу тәжірибесінің молдығы. Заңдылықтың ең бір маңызды деп саналатын тағы бір жағы – ол мәтін бөлігінің көлемін анықтау мен статистикалық баға беруде (сынауда) қолдану мүмкіндігі.

Айнымалы және үздіксіз x шама үлестірілудің $\varphi(x)$ функцияның аргументі ретінде $-\infty$ *(минус шексіздік)* пен $+\infty$ *(плюс шексіздік)* аралығындағы шамаларды қабылдайды. Ал *"x"* шаманың үлестірілуінің нормальды заңдылығы төмендегі пішінде көрініс табады:

$$n^m = \varphi(x) \cdot n = \frac{n \cdot \Delta x}{\sigma} \cdot \frac{1}{\sqrt{2\Pi}} \cdot \ell^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (5)$$

Мұндағы Δx – дискретті вариациялық қатардан үздіксіз қатарға ауысуға қажетті интервал ұзындығы, a – "математикалық күту" (жоғарыдағы 4-ші өрнекпен анықталады). Ал σ^2 – дисперсия немесе ол *"екінші реттегі орталық момент"* деп те аталады және оның математикалық өрнегі төмендегідей:

$$\sigma^2 = \mu_2 = \frac{1}{n-1} \cdot \sum_i x_i \cdot (x_i - a)^2. \quad (6)$$

Математикалық күту *(a)* (4) мен дисперсия (σ^2) (6) статистикалық параметрлер арқылы үлестірілу заңдылығының маңызды жақтары сипатталады. Мысалы, олар арқылы өзгермелі x_i мәндерінің шашырамдық сипаты мен сызбадағы үлестірілудің қисық сызықтық пішінінің орналасу қалпын анықтауға болады. Ал тәжірибе жүзінде алынған үлестірілу сипатының теориялық "нормаль заңдылығына" үйлесу дәрежесін анықтау (сәйкестігін бағалау) үшін "ассиметрия" және "эксцесс" параметрлерін білу қажет деп саналады:

$$A = \frac{\mu_3}{\sigma^3}, \quad (7)$$

$$E = \frac{\mu_4}{\sigma^4} - 3. \quad (8)$$

Мұндағы σ – орта квадраттық ауытқу, μ_3 – мен μ_4 – үшінші және төртінші реттегі орталық моменттер:

$$\mu_3 = \frac{1}{n} \cdot \sum_i n_i \cdot (x_i - a)^3, \quad (9)$$

$$\mu_4 = \frac{1}{n} \cdot \sum_i n_i \cdot (x_i - a)^4. \quad (10)$$

Нормаль үлестірілу заңдылығы үшін μ_3 мен μ_4 мәндері нөлге тең болады. Сондықтан сынаққа түсіп отырған үлестірілу заңдылығы үшін "ассиметрия" *(A)* мен "эксцесс" *(E)* мәндері де нөлге жуық келетін шама болса, онда тәжірибелік үлестірілу *нормаль* заңына жақын деп жорамалдауға әбден болады. Керісінше, *A* мен *E* шамаларының нөлден алшақтығы өскен сайын, сынаққа түскен тәжірибелік үлестірілудің де *нормаль* заңдылығына үйлесімділігі алшақтай түседі.

Теориялық үлестірілу заңдылығының *Шарлье A, B* мен *логарифмдік нормаль* түрлері жайлы толығырақ мына әдебиеттерден танысуға болады: [80, 94].

3.5.3. Пирсонның χ^2 (хи квадрат) үйлесімдік критерийі

Жоғарыда сөз болған *"Пуассон"* мен *"нормаль"* және *"Шарлье A, B"* үлестірілу заңдылықтары *теориялық заңдылық* немесе *ықтималдық модель (үлгі)* деп аталады. Зерттеуге алынған тілдік бірліктің тәжірибе жүзінде анықталған тәжірибелік (эмпирикалық) үлестірілулердің жоғарыда сөз болған кейбір теориялық заңдылықтарға үйлесімділігі тілдік бірліктің мәтін бойында заңды таралуының айғағы деуге болады. Сондықтан осындай үйлесімділіктің (жақындықтың) дәрежесін "бағалайтын" мүмкіндіктің қажеттігі туады. Мұндай мүмкіндікке статистикалық лингвистика саласында бұрыннан да қолданылып жүрген Пирсонның χ^2 (хи квадрат) үйлесімдік критерийі (критерий согласия) жатады [15, 16, 93]. Ол үшін таңдама бөліктің эмпирикалық үлестірілуі бас жиынның теориялық үлестірілуіне сәйкес келеді деген "нөлдік" гипотеза (ғылыми болжам) – H_0 ұсынылуы қажет. Ол болжамды (гипотезаны) "қабылдау" немесе "терістеу" үшін тәжірибелік (эмпирикалық) пен теориялық үлестірулер арасындағы алшақтықты анықтайтын мынадай статистикалық сипаттама құрастырылады:

$$W=\sum_{i=i}^{k}\frac{(n_i^{ЭМ}-n_i^T)^2}{n_i^T}. \qquad (11)$$

Мұндағы $n_i^{ЭМ}$ және n_i^T – мәтіндегі лингвистикалық бірліктің тәжірибелік және теориялық жиіліктері. Егер мәтін көлемі *(N)* шексіздікке ұмтылса, яғни $N\to\infty$, онда статистикалық сипаттама *W*-ның χ^2 (хи квадрат) үлестірілу заңдылығын қанағаттандыратынына көз жеткізуге болады. Ол үлестірілудің еркіндік дәреже саны $\nu=t-\ell-1$. Мұндағы ℓ – сызықтық қатынаста тұратын параметрлер саны, t – тәжірибелік үлестірілу тобының саны.

Сонымен, тәжірибелік (эмпирикалық) және теориялық үлестірілулердің сәйкестігін эмпирикалық жиіліктердің ($n_i^{ЭМ}$) математикалық күтуінен ($n_i^T=NP_i$) алшақтық дәрежесі бойынша бағалауға болады (p_i – ықтималдық, N – таңдама бөліктің көлемі).

Келесі математикалық өрнек χ^2 (хи квадрат) критерийі деп аталады:

$$\chi^2 = \sum_{i=i}^{k} \frac{(n_i^{ЭМ} - n_i^T)^2}{N \cdot P_i} . \qquad (12)$$

Егер есептеліп шығарылатын $\chi^2=0$ болса, онда эпирикалық (тәжірибелік) және теориялық жиіліктер біріне-бірі сәйкес келеді деп ұйғарылады. Ал басқа жағдайларда, яғни χ^2 мәні нөлден айырықша болған жағдайда, аталған жиіліктердің сәйкестік сипатының әлсіреуі χ^2 мәнінің нөлден алшақтығына тура пропорционал деп есептеледі. Дәлірек айтқанда, χ^2 мәнінің нөлден алшақтығы өскен сайын, тәжірибелік және теориялық жиіліктер сәйкестігі соншалықты әлсірей түседі. Осымен бірге, алшақтық сипатының мәнді (существенное) және мәнсіз (несущественное) түрлері де ажыратылады. Мұндай жайт χ^2 мәні мен теориялық үлестірілудің басқа да параметрлері толық анықталғаннан кейін ғана белгілі болады. Басқаша айтқанда, тәжірибелік және теориялық жиіліктер арасындағы сәйкестік болжамын (H_0) χ^2 критерийі бағалай алады. Ол үшін "еркіндік дәреже санын" (ν) анықтау керек және осы критерий үшін пайыздық мағыналық деңгей" (уровень значимости) – q таңдалып алыну қажет. Мысалы, ықтималдықтың 0,95 мәніне мағыналық деңгейдің $q=0,05$ немесе 5 пайыздық деңгейі сәйкес келеді. Бұл шама алдын ала қабылданған шекаралық аядан сырт шығуы кездейсоқтық жағдайға ғана тән екендігін және оның ықтималдығы $q=0,05$ тең немесе мұндай жағдайдың 5 пайызы ғана шындықтан тыс жатқанын білдіреді. Сонымен бірге зерттеуші арнайы кесте арқылы χ_q^2 мәнін анықтай білуі керек. Мұндағы χ^2, еркіндік саны ν мен мағыналық деңгейге (q) тәуелді, яғни аргументтері ν және q болатын функция: $\chi_q^2=\chi^2(\nu, q)$; χ^2-тің есептеліп шыққан мәні мен оның кестелік мәнін (χ_q^2) салыстыру қажеттігі туындайды. Бұндай салыстыру кезінде екі түрлі теңсіздіктердің кездесуі мүмкін: $\chi^2 \geq \chi_q^2$ немесе $\chi^2 < \chi_q^2$. Бірінші жағдайда, яғни $\chi^2 \geq \chi_q^2$ болса, критерий мәні алшақтық дәрежесінің мәнділік аясына қатысты болады да тәжірибелік пен теориялық үлестірілулердің сәйкестігі жайлы H_0 болжам теріске шығарылып, қабылданбауы керек. Ал, екіншіде, яғни $\chi^2 < \chi_q^2$ болса, аталған үлестірілулердің арасындағы алшақтық – кездейсоқ (мәнсіз) жағдайға жатқызылып, H_0 болжам қабылдануы қажет.

H_0 болжамды "бағалау" осындай салыстырулардан басқа да жолмен іске асатынын айта кетейік. Ол әдіс $p(\chi^2 \geq \chi_q^2)$ ықтималдық шамасын білуді қажет етеді. Егер осындай ықтималдық алдын ала берілген мағыналық деңгейден, мысалы, $q=0,05$ шамадан төмен жатса, онда ол кездейсоқ оқиға ретіндегі алшақтықтардың ықтималдығы өте аз екендігін білдіреді. Бұл жағдайда тәжірибелік пен теориялық үлестірілулердің арасындағы алшақтық кездейсоқ емес (мәнді), сондықтан сәйкестікті қолдайтын H_0 болжам қабылданбайды, ол – "жалған" деген сөз.

Керісінше, егер $p(\chi^2 \geq \chi_q^2)$ ықтималдық шамасы $q=0,05$ санынан айтарлықтай үлкен болса, онда аталған екі үлестірілулер арасы жақын жатыр, сондықтан H_0 гипотезасы шындыққа сай келеді және ол қабылданады.

Сонымен, $p(\chi^2 \geq \chi_q^2)$ ықтималдық мәні H_0 болжамының (гипотезаның) қабылдану-қабылданбауына бағалық критерийі бола алады.

Математикалық статистика саласының белгілі маманы Т.Крамер бірнеше тәжірибелер нәтижелерін талдай келе, мынадай қорытындыға келеді: "Тәжірибелік және теориялық үлестірілулердің сәйкестігі "жақсы" деп бағаланады, егер $p(\chi^2 \geq \chi_q^2)=0,17$ болса, ал осы ықтималдықтың $0,91$-ге тең мәнінде сәйкестік "өте жақсы" деп бағалануы керек. Екі үлестірілулер арасындағы алшақтық "мәнді дәрежеде" қалуы үшін $p(\chi^2 \geq \chi_q^2)=0,03$ болуы жеткілікті" [81], – дейді.

Енді, осындай зерттеуді, яғни эмпирикалық үлестірілудің қалап алынған теориялық заңдылықпен сәйкестігінің мәнді не мәнсіз екендігін анықтау үшін жүргізілетін тәжірибе реті төмендегідей [31]:

1. Зерттеуге тиісті мәтінді және тілдік бірлікті таңдау. Мысалы, қазақ көркем әдебиет мәтіндеріндегі зат есім сөз табына қатысты сөздер.

2. $B_N^n = k$ көрінісіндегі таңдалып алынған параметрлеріне (N, n, k) сай етіп, зерттелетін мәтінді таңдама бөліктерге бөлу.

3. N, n, k параметріне және тілдік бірлікке сәйкес жиілік сөздіктер түзу.

4. Жиілік сөздіктер негізінде дискретті не үздіксіз вариациялық қатар құру.

5. Тиісті математикалық өрнекті (формуланы) пайдаланып, жоғарыда сөз болған мына параметрлерді есептеп шығару:

а) математикалық күтілуді (a);

ә) ауытқудың квадраттық ортасын (σ);

б) ассиметрия (A) мен эксцесті (E).

в) тілдік бірліктің теориялық жиілігін $(n_i^T = NP_i)$;

г) әрбір "i" тобына қатысты "хи квадрат" (χ^2) мәнін.

6. Жиіліктің *5-тен* кіші мәндерін қажеттілікке сай біріктіру және $\sum_{i=1}^{k} \chi_i^2$ қосындысына қажетті өзгертулер енгізу (қолмен).

7. Еркіндік дәреже санын анықтау. Мысалы, үлестірілімнің Пуассон заңы үшін $\ell = 1$, нормаль мен логнормаль, Шарлье В заңдары үшін $\ell = 2$, Шарлье А үшін $\ell = 4$.

8. Есептеліп шығарылған χ^2 пен еркіндік дәреже саны (V) бойынша $p(\chi^2 \geq \chi_q^2)$ – ықтималдық интегралдың кестелік мәнін табу [81].

Тілдік бірліктің мәтін ішіндегі үлестірілу сипатын анықтау сериялар саны *(k)* мен таңдама бөліктер көлемін *(n)* дұрыс нормалауға байланысты болады. Себебі, көлемі *N*-ге тең бір мәтінді саны *k*-дан тұратын сериялларға әртүрлі жолмен бөлуге болады. Ол ішкі таңдама мәтін бөліктерінің көлеміне *(n)* байланысты. Мысалы, мәтін көлемі *N=5000* сөзқолданыстан тұрса, ал ішкі бөліктер көлемі *n=100* болса, серия саны *k=50; дәл* сол сияқты *n=500* болса, *k=5* болады. Әр уақытта $N=k\cdot n$ және *k=N:n*, яғни $N=5000=50\cdot 100$ және *k=5000:100=50;* $N=5000=10\cdot 500$; $N=5000=5\cdot 1000$ және т.б.

Бұл жерде ескерте кететін жайт, лингвистикалық зерттеулерде сериялар саны мәтіндегі сөзқолданыстар санынан басқа, кітаптың жол не бет санымен де өлшенуі мүмкін. Сол сияқты, параметрлерді нормалау кезінде мәтін бірліктерінің қайсысы және қандай көлемде алынатыны алдын ала нақтылануы қажет. Әрбір нормалау сипатына сай және өзіне ғана тән теориялық үлестірілу заңдылығы сәйкес келуі мүмкін.

Мәселен, нақты бір нормалауда сынаққа алынған тілдік бірлік "Пуассон үлестірілу" заңына, ал басқаша түрдегі нормалауда сол бірлік мәтін ішінде "Нормаль үлестірілу" заңына "бағынуы" немесе белгілі бір ішкі бөлік көлемінен бастап тек бір ғана теориялық заңдылыққа сәйкес келуі мүмкін екені анықталды [15, 63, 88]. Сондықтан үлестірілу заңдылықтарының параметрлерін салыстыруда және оған лингвистикалық тұрғыда түсініктеме беруде зерттеу нысанын (мәтінді) нормалау сипатына аса көңіл бөлінуі қажет.

Сонымен, мәтінді ішкі бөліктеудің ең тиімді жолын табу лингвистикалық модельдеу ісінде маңызды деп саналады.

Енді қазақ мәтіндеріне қатысты жүргізілген кейбір тәжірибе нәтижелеріне тоқталайық [31].

Сөз таптарының қолдану жиілігін сөз еткенде, барлық тілдерге де ортақ заңдылық – зат есім мен етістік сөздердің өнімді қолданылуы. Мысалы, М.Әуезовтің "Абай жолы" романы мәтінінде зат есім 35,36%, ал етістік сөздер 31,44%, яғни 100 пайыздың 70-не жуығы тек осы екі сөз табына қатысты сөздер екен. Газет мәтінінің 45% – зат есім, 22 пайызы – етістік (Ахабаев). Сондықтан қазақ мәтіндеріндегі сөз үлестірілу заңдылығын зерттеу кезінде әр сөз табына байланысты алынатын таңдама мәтіндер осындай қарапайым және тәуелсіз талаптарға жауап берулері керек.

Егер аталған шарттар орындалды деп ұйғарылса, зат есім, етістік сөздердің (не басқа сөз таптарының) тәжірибелік (эмпирикалық) үлестірілудің теориялық (нормальды не басқа) заңдылыққа үйлесімді келеді (согласуется) деп *"H_0"* жорамалды жасауға болады. Ал жорамалды "бағалау" әртүрлі *N, n, k* шамаларына қатысты $p(\chi^2 \geq \chi^2_q)$ ықтималдылықтарды салыстыру негізінде іске асады.

Біздің тәжірибемізде "Абай жолы" романынан алынған *N=50000* сөзқолданыс мынадай нормалауға сәйкес келді:

а) $N_1=2500$, $N_2=5000$, $N_3=10000$, $N_4=20000$, $N_5=25000$, $N_6=40000$ және $N_7=50000$ көлемдері ретінде бөлек-бөлек қарастырылды;

ә) ішкі бөлік көлемі *(n)*: $n_1=100$, $n_2=200$, $n_3=250$, $n_4=500$, $n_5=1000$ сөзқолданыстан тұратын көлемдері ретінде бөлек-бөлек қарастырылды;

б) барлық жағдайда $N=k\times n$ болатынын ескеріп, сериялар шамасы *k=N:n* анықталды: $k_1=25$, $k_2=50$, $k_3=100$, $k_4=200$, $k_5=500$ сериялық сан мөлшері екені анықталды.

Осылайша нормалаудың нәтижесінде көркем әдебиет мәтіні бойынша зат есім сөздердің тәжірибелік жиіліктері оның теориялық жиіліктерімен сәйкес келудің "сенімділік дәрежесін" қанағаттандыру жағын алғанда 1-ші орында "Нормаль үлестірілу" заңы, ал 2-ші орында "Шарлье А үлестірілу" заң түрі екені анықталды. Газет мәтіні бойынша, сериялар саны 100-ден артқан жағдайда зат есім сөздердің теориялық заңдылықтарға сәйкес келу мүмкіндігі "логнормаль үлестірілу" заңдылығы екені дәлелденді *(3.9-3.12-кестелерді* қараңыз).

Көркем әдебиет және публицистика мәтіндеріндегі етістік сөздердің тәжірибелік үлестірілу деректерінің теориялық үлестірілу заңдарымен үйлесімділігін зерттей келе, олардың математикалық моделі (үлгісі) нормаль заңдылығы екені айқындалды. Ал егер, серия саны $k \leq 100$ болған жағдайда, етістік сөздердің үлестірілулері теориялық заңдылықтар ішінен "Шарлье А үлестірілу" типіне сәйкестігі айрықша байқалды.

Публицистикалық мәтіндердегі етістікке қатысты сөздердің тәжірибелік үлестірілуі параметрлерді нормалаудың мынандай түрінде: $\{N \geq 1000,\ n \geq 100,\ k \geq 100\}$, теориялық заңдылықтар ішінен "нормаль үлестірілуін" көбірек қанағаттандыратыны *3.13* және *3.14-кестелердегі* ықтималдықтың дәрежелерінен байқауға болады.

Сонымен, қазақ тіліндегі етістікке қатысты сөздердің тәжірибелік жиіліктерінің мәтін бойындағы үлестірілуі (аталған нормалауға сай) "Нормаль заңымен" үйлесімдік табады деп қорытындылауға болады.

Теориялық үлестірілу заңдарының Пуассон және Шарлье В типі қазақ тіліндегі *зат есім* мен *етістік* сөздердің тәжірибелік жиілік үлестірілу деректерімен үйлеспейтіндігі де байқалды.

Зат есім мен *етістік* сөздердің қатарында *сын есім* сөздердің де мәтіндегі үлестірілу заңдылығы қарастырылды. Қысқаша айтқанда, олардың теориялық заңдылықтарымен сәйкестігі көркем әдебиет пен газет мәтіндерінде бірдей емес және нормалау түріне қарай, яғни N, n, k параметрлерін бөлшектеуге қатысты айырым табады. Дәлірек айтқанда, көркем әдебиет мәтіні үшін нормалау сипаты – $\{N \geq 2500,\ n \geq 1000,\ k \geq 25\}$ болғанда, "Пуассон үлестірілу" мен "Шарлье үлестірілу" заңдылығының А мен В типтерімен бірдей дәрежеде үйлесетіні байқалды, ал газет тіліне қатысты $\{N \geq 5000,\ n \geq 100,\ k \geq 50\}$ нормалау әдісі ғана аталған заңдылықтарды қанағаттандырады екен.

Үйлесімдік дәрежесін оның "артықшылық" тұрғысынан қарастырсақ, әрине, *сын есім* сөздерге қатысты Пуассонның теориялық үлестірілу заңы бірінші орынға ие. Солай бола тұра, нормалаудың $\{N \geq 2000,\ n \geq 100,\ k \geq 200\}$ мәндерінен бастап, қазақ тілінің *сын есім* сөздері "нормаль заңдылығымен" де жақсы үйлесімдік табатынын байқаймыз *(3.15, 3.16-кестелер)*. Ал *3.17-кестеде "үстеу"* сөздердің 5 түрлі теориялық заңдылықтармен үйлесімдігін аңғартатын $p(\chi^2 \geq \chi_q^2)$ шамалары көрініс тапты.

Сөз таптарының тәжірибелік жиіліктерінің теориялық үлестірілу заңдылықтарымен үйлесімдігі, көбіне, сериялар санына қатысты екені байқалады. *Үстеу* сөздерге қатысты ең "қолайлы" деп Пуассон үлестірілу заңын атауға болады. Ал сериялар саны $k>500$ болғанда үстеу сөздер "нормаль заңымен" үйлеседі.

Сонымен, жоғарыда баяндаған жайттар мәтінді мөлшерлеу (нормалау) әдістерінің қазақ тіліндегі сөз таптарының тәжірибелік (эмпирикалық) үлестірілулерінің белгілі бір теориялық үлестірілу заңдылығына сәйкес келу не келмеу мүмкіндіктерін айқындауды көздейді. Бұл зерттеудің нәтижелері қазақ тіліндегі құбылыстарды модельдеу мәселелерінде аса маңызды рөл атқаратыны сөзсіз.

3.9-кесте

Көркем әдебиет мәтініндегі *зат есім* сөздер үлестірілуінің $p(\chi^2 \geq \chi_q^2)$ мәндері

Таңдама нормасы *B(N/n)=k*	Үлестірілу заңдары		
	Нормаль	Шарлье А	Логнормаль
B(2500/100)=25	*0,660*	*0,746*	*0,175*
B(5000/200)=25	*0,734*	*0,548*	*0,006*
B(10000/400)=25	*0,862*	*0,713*	*0,0003*
B(5000/100)=50	*0,991*	*0,972*	*0,368*
B(10000/200)=50	*0,625*	*0,380*	*0,246*
B(10000/100)=100	*0,101*	*0,193*	*0,053*

3.10-кесте

Публицистика мәтініндегі *зат есім* сөздер үлестірілуінің $p(\chi^2 \geq \chi_q^2)$ мәндері (*k* – тұрақты)

Таңдама нормасы *B(N/n)=k*	Үлестірілу заңдары		
	Нормаль	Шарлье А	Логнормаль
B(2500/100)=25	*0,797*	*0,994*	*0,999*
B(5000/200)=25	*0,986*	*0,935*	*0,231*
B(12500/500)=25	*0,740*	*0,486*	*0,005*
B(25000/1000)=25	*0,979*	*0,880*	*0,000*
B(5000/100)=50	*0,687*	*0,781*	*0,619*
B(10000/200)=50	*0,302*	*0,118*	*0,064*
B(25000/500)=50	*0,775*	*0,398*	*0,692*
B(50000/1000)=50	*0,548*	*0,141*	*0,0003*
B(10000/100)=100	*0,626*	*0,563*	*0,839*
B(20000/200)=100	*0,983*	*0,976*	*0,750*
B(50000/500)=100	*0,991*	*0,990*	*0,413*
B(20000/100)=200	*0,310*	*0,456*	*0,477*
B(40000/200)=200	*0,575*	*0,658*	*0,677*
B(50000/250)=200	*0,695*	*0,699*	*0,756*
B(50000/100)=500	*0,802*	*0,649*	*0,809*

3.11-кесте

Публицистика мәтініндегі *зат есім* сөздер үлестірілуінің $p(\chi^2 \geq \chi_q^2)$ мәндері (*n* – тұрақты)

Таңдама нормасы *B(N/n)=k*	Үлестірілу заңдары		
	Нормаль	Шарлье А	Логнормаль
B(2500/100)=25	*0,797*	*0,994*	*0,999*
B(5000/100)=50	*0,687*	*0,781*	*0,619*
B(10000/100)=100	*0,626*	*0,563*	*0,839*
B(20000/100)=200	*0,310,*	*0,456*	*0,477*
B(50000/100)=500	*0,802*	*0,649*	*0,809*
B(5000/200)=25	*0,986*	*0,936*	*0,231*
B(10000/200)=50	*0,302*	*0,118*	*0,064*
B(20000/200)=100	*0,983*	*0,976*	*0,750*
B(40000/200)=200	*0,575*	*0,658*	*0,677*
B(12500/500)=25	*0,740*	*0,486*	*0,005*
B(25000/500)=50	*0,775*	*0,398*	*0,692*
B(50000/500)=100	*0,991*	*0,990*	*0,413*
B(25000/1000)=25	*0,978*	*0,880*	*0,000*
B(50000/1000)=50	*0,548*	*0,141*	*0,0003*
B(50000/250)=200	*0,695*	*0,699*	*0,756*

3.12-кесте

Публицистика мәтініндегі *зат есім* сөздер үлестірілуінің $p(\chi^2 \geq \chi_q^2)$ мәндері (*N* – тұрақты)

Таңдама нормасы *B(N/n)=k*	Үлестірілу заңдары		
	Нормаль	Шарлье А	Логнормаль
B(5000/200)=25	*0,986*	*0,936*	*0,231*
B(5000/100)=50	*0,687*	*0,781*	*0,619*
B(25000/1000)=25	*0,979*	*0,880*	*0,000*
B(25000/500)=50	*0,775*	*0,398*	*0,692*
B(20000/200)=100	*0,983*	*0,976*	*0,750*
B(20000/100)=200	*0,310*	*0,456*	*0,477*
B(50000/1000)=50	*0,548*	*0,141*	*0,0003*
B(50000/500)=100	*0,991*	*0,990*	*0,413*
B(50000/250)=200	*0,695*	*0,699*	*0,756*
B(50000/100)=500	*0,802*	*0,649*	*0,809*
B(40000/200)=200	*0,575*	*0,658*	*0,677*
B(2500/100)=25	*0,797*	*0,994*	*0,999*
B(12500/500)=25	*0,740*	*0,486*	*0,005*
B(10000/200)=50	*0,302*	*0,118*	*0,064*
B(10000/100)=100	*0,626*	*0,563*	*0,839*

3.13-кесте

Көркем әдебиет мәтініндегі *етістік* сөздер үлестірілуінің $p(\chi^2 \geq \chi_q^2)$ мәндері

Таңдама нормасы *B(N/n)=k*	Үлестірілу заңдары		
	Нормаль	Шарлье А	Логнормаль
B(2500/100)=25	*0,603*	*0,544*	*0,395*
B(5000/200)=25	*0,530*	*0,788*	*0,539*
B(10000/400)=25	*0,156*	*0,807*	*0,393*
B(10000/200)=50	*0,319*	*0,547*	*0,679*
B(10000/100)=100	*0,320*	*0,350*	*0,495*

3.14-кесте

Публицистика мәтініндегі *етістік* сөздер үлестірілуінің $p(\chi^2 \geq \chi_q^2)$ мәндері

Таңдама нормасы *B(N/n)=k*	Үлестірілу заңдары		
	Нормаль	Шарлье А	Логнормаль
B(2500/100)=25	*0,151*	*0,080*	*0,224*
B(5000/200)=25	*0,978*	*0,978*	*0,009*
B(25000/1000)=25	*0,045*	*0,074*	*0,000*
B(5000/100)=50	*0,469*	*0,213*	*0,005*
B(10000/200)=50	*0,187*	*0,609*	*0,614*
B(25000/500)=50	*0,174*	*0,369*	*0,0003*
B(10000/100)=100	*0,926*	*0,729*	*0,140*
B(20000/200)=100	*0,187*	*0,152*	*0,072*
B(50000/500)=100	*0,238*	*0,048*	*0,100*
B(40000/200)=200	*0,871*	*0,587*	*0,285*
B(50000/250)=200	*0,855*	*0,636*	*0,067*
B(50000/100)=500	*0,302*	*0,109*	*0,253*

3.15-кесте

Көркем әдебиет мәтініндегі *сын есім* сөздер үлестірілуінің $p(\chi^2 \geq \chi_q^2)$ мәндері

Таңдама нормасы	Үлестірілу заңдары				
	Норм.	Шар.-А	Лог-нор	Пуас.	Шар.-В
B(2500/100)=25	*0,215*	*0,019*	*0,078*	*0,024*	*0,044*
B(5000/200)=25	*0,000*	*0,013*	*0,000*	*0,014*	*0,005*
B(10000/400)=25	*0,000*	*0,040*	*0,000*	*0,133*	*0,056*
B(5000/100)=50	*0,384*	*0,623*	*0,141*	*0,848*	*0,749*
B(10000/200)=50	*0,075*	*0,401*	*0,024*	*0,529*	*0,477*
B(10000/100)=100	*0,93*	*0,925*	*0,661*	*0,674*	*0,548*

3.16-кесте

Публицистика мәтініндегі *сын есім* сөздер үлестірілуі [$p(\chi^2 \geq \chi_q^2)$]

Таңдама нормасы	Үлестірілу заңдары				
	Норм.	Шар.-А	Лог-нор.	Пуас.	Шар.-В
B(2500/100)=25	*0,287*	*0,216*	*0,253*	*0,331*	*0,121*
B(5000/200)=25	*0,966*	*0,999*	*0,000*	*0,851*	*0,989*
B(12500/500)=25	*0,055*	*0,339*	*0,000*	*0,410*	*0,317*
B(25000/1000)=25	*0,013*	*0,469*	*0,000*	*0,114*	*0,433*
B(50000/2000)=25	*0,000*	*0,269*	*0,000*	*0,401*	*0,819*
B(5000/100)=50	*0,227*	*0,277*	*0,012*	*0,188*	*0,079*
B(10000/200)=50	*0,333*	*0,644*	*0,148*	*0,690*	*0,734*
B(12500/250)=50	*0,602*	*0,779*	*0,130*	*0,481*	*0,420*
B(25000/500)=50	*0,002*	*0,652*	*0,038*	*0,328*	*0,211*
B(50000/1000)=50	*0,005*	*0,572*	*0,000*	*0,282*	*0,175*
B(10000/100)=100	*0,017*	*0,054*	*0,172*	*0,104*	*0,159*
B(20000/200)=100	*0,047*	*0,0002*	*0,162*	*0,065*	*0,123*
B(25000/250)=100	*0,181*	*0,172*	*0,137*	*0,050*	*0,071*
B(20000/100)=200	*0,965*	*0,874*	*0,760*	*0,450*	*0,407*
B(40000/200)=200	*0,211*	*0,040*	*0,099*	*0,201*	*0,280*
B(50000/250)=200	*0,253*	*0,083*	*0,021*	*0,130*	*0,070*
B(50000/100)=500	*0,101*	*0,132*	*0,221*	*0,308*	*0,340*

3.17-кесте

Публицистика мәтініндегі *үстеу* сөздер үлестірілуі [$p(\chi^2 \geq \chi_q^2)$]

Таңдама нормасы	Үлестірілу заңдары				
	Норм.	Шар.-А	Лог-нор.	Пуас	Шар-В
B(2500/100)=25	*0,000*	*0,000*	*0,00001*	*0,105*	*0,337*
B(5000/200)=25	*0,000*	*0,00001*	*0,00003*	*0,535*	*0,699*
B(12500/500)=25	*0,000*	*0,000*	*0,000*	*0,475*	*0,798*
B(25000/1000)=25	*0,000*	*0,000*	*0,000*	*0,999*	*0,996*
B(50000/2000)=25	*0,000*	*0,000*	*0,000*	*0,995*	*0,984*
B(5000/100)=50	*0,010*	*0,0004*	*0,002*	*0,897*	*0,867*
B(10000/200)=50	*0, 050*	*0,199*	*0,0004*	*0,907*	*0,855*
B(12500/250)=50	*0,070*	*0,067*	*0,004*	*0,9992*	*0,997*
B(25000/500)=50	*0,001*	*0,045*	*0,005*	*0,863*	*0,902*
B(50000/1000)=50	*0,042*	*0,046*	*0,042*	*0,600*	*0,963*
B(10000/100)=100	*0,863*	*0,866*	*0,042*	*0,401*	*0,645*
B(20000/200)=100	*0,388*	*0,128*	*0,043*	*0,364*	*0,532*
B(50000/500)=100	*0,174*	*0,780*	*0,147*	*0,576*	*0,910*
B(20000/100)=200	*0,186*	*0,069*	*0,015*	*0,450*	*0,448*
B(40000/200)=200	*0,386*	*0,742*	*0,277*	*0,433*	*0,537*
B(50000/200)=250	*0,475*	*0,667*	*0,223*	*0,486*	*0,542*
B(25000/50)=500	*0,504*	*0,490*	*0,208*	*0,396*	*0,567*

3.6. Колмогоровтың үйлесімдік критерийі арқылы лингвистикалық болжамды сынау (тексеру)

Мәтінді сипаттаудың аналитикалық тұрпаты мен соған жуықтайтын моделін (үлгісін) табудың қолданбалы және теориялық маңыздылығы зор. Тіл білімінде алғаш рет орыс тілі мәтініне қатысты сөзтұлғаның жиілік сөздіктегі орнын анықтаудың ықтималдық үлестірілуін сипаттауға арналған Г.Г.Белоноговтың зерттеу жұмысын атауға болады [20]. Өз зерттеуінде ғалым Вейбуллдың формуласын пайдаланады [96]:

$$F(x)=1-e^{-cx^{k}}. \qquad (1)$$

Мұндағы x – сөзтұлғаның сөздіктегі рет саны, "c мен k" – Вейбулл үлестірілуінің параметрлері, "e" – натурал логарифмнің негізі. Осы (1) өрнекпен анықталатын $F(x)$ функцияның мәні – мәтін ішінде кездесетін сөзтұлғаның сол мәтіннен түзілген жиілік сөздіктегі кез келген жиі кездесетін саны "x"-ке тең сөзтұлғалармен сәйкес келу ықтималдығын білдіреді. Ғалымдар, көп жағдайда, зерттеу нысаны ретінде орыс, ағылшын, неміс тілдерінің мәтіндерін қарастырып, аталған өрнекті сөзтұлғаның немесе сөз тіркестің "теориялық қатынастық жиіліктер жиынтығын" табу үшін қолданады. Ал біздің ізденісімізде ол өрнек тұңғыш рет қазақ тіліндегі көркем әдебиет мәтіндеріндегі сөзтұлғаларды сипаттауда пайдаланылды.

Енді аталған (1) өрнекті басқаша тұрпатта жазайық. Ол үшін $F(x)$ үлестірілу функциясының орнына теориялық қатынастық жиіліктердің жиынтығын (қосындысын) – f_i^{*Teop} аламыз.

Сонда ол $f_i^{*Teop}=1-e^{-cx^{k}}$ өрнегі арқылы жазылады.

Мұндағы $x=i$ – жиілік сөздіктегі сөзтұлғаның рет саны, $f_i^{*Teop} < 1$, ал $e^{-ci^{k}} \geq 0$ болғандықтан соңғы теңдіктің екі жағын логарифмдеуге мүмкіндік туады:

$$-ci^{k}\ell ne=\ell n(1-f_i^{*Teop})$$

немесе

$$-ci^{k}=-\ell n(1-f_i^{*Teop}). \qquad (2)$$

Енді (2) теңдіктегі $[-\ell n(1-f_i^{*Teop})]=y$ және $i=x$ деп белгілесек, аталған өрнек мынадай пішінді қабылдайды: $Cx^{k}=Y$, мұндағы $Y \geq 0$, сондықтан теңдіктің екі жағынан бірдей логарифмдей аламыз: $lnC+klnx=lny$.

Егер $lnC=A$, $lnx=X$, $lny=Y$ деп белгілесек:

$$A+kX=Y. \qquad (3)$$

Әрі қарайғы әрекетімізде "ең кіші квадраттар әдісі" (метод наименьших квадратов) бойынша A мен k мәндерін анықтап, C-ны табуға болады: [20, 47]:

$$C=l^{A}. \qquad (4)$$

Әрбір рет саны "*i*"-ге сай келетін *k* мен *C*-ны анықтап, оған сәйкес келетін f_i^{*Teop} табылады.

Сонымен, әрбір *i*-ге қатысты тәжірибелік $f_i^{*эмп}$ және теориялық $f_i^{*Teop.}$ қатынастық жиіліктердің жиынтықтары арқылы "орташа квадраттық ауытқуды" мынандай өрнекпен есептеп шығаруға болады:

$$\sigma=\sqrt{\frac{1}{q}\sum_i (f_i^{*эмп.}-f_i^{*Teop.})^2}\ , \qquad (5)$$

мұндағы q – тәжірибе саны.

Жоғарыда баяндалған жайттардың негізінде әрбір рет саны *i*-ге қатысты *C, k,* σ мәндерін арнайы компьютерлік бағдарлама арқылы есептеп шығаруға болады [24]. Осындай бағдарламаның негізінде М.Әуезовтің "Абай жолы" романының 5 түрлі жиілік сөздіктерінің соңғы рет санына $i_{соң}$ қатысты аталған параметрлердің мәндері *3.18-кестеде* көрсетілді.

Бұл кестедегі "орта квадраттық ауытқу" (σ) мәндері кездейсоқ шаманың "математикалық күтілу" *(a)* (математическое ожидание) аясында таралуын (рассеивание) сипаттайды. Кестедегі деректерге көңіл бөлсек, олардың таралу сипатының аздығы байқалады.

3.18-кесте

Параметрлер	Сөздіктер				
	1-кітап	2-кітап	3-кітап	4-кітап	5-роман
$i_{соң}$	3276	3872	3570	4027	9614
C	0,0452	0,0429	0,0425	0,0374	0,0496
K	0,4196	0,4174	0,4150	0,4318	0,3876
σ	0,0064	0,0057	0,0061	0,0060	0,0050

Бағдарлама көмегімен әр сөздікке қатысты есептелген $f_i^{*Teop.}$ мен тәжірибелік $f_i^{*эмп}$ мәндерін жай ғана (көзбен) салыстырғанның өзінде көп ұқсастықты байқауға болады. Бірақ бұлай салыстыру нәтижелері әр уақытта дұрыс бола бермеуі мүмкін. Сондықтан оған математикалық баға берілуі тиіс. Осындай қарапайым құрал ретінде біз "Колмогоровтың үйлесімдік критерийін" пайдалануды жөн санадық [24, 80, 81]. Аталған критерий статолингвистика саласындағы теориялық үлестірілуді таңдау жағдаятында жиі қолданылып жүр.

Біздің зерттеу барысындағы алға қойылған мақсат – "Абай жолы" романы мәтінінің сөзтұлға жиіліктерінің үлестірілуі H_0 жорамалы бойынша, кездейсоқ оқиғаның *(x)* алдын ала берілген *F(x)* функциясының үлестірілуімен үйлесетіндігін тексеру (бағалау).

А.Н.Колмогоров тәсілі бойынша, теориялық пен тәжірибелік үлестірілулер арасындағы үйлесімдік (не үйлеспеушілік) өлшемі ретінде үлестірілудің эмпирикалық $f^{*эмп}(x)$ пен оған сәйкес келетін теориялық функция $F(x)$ айырым модулінің максимумдық шамасы алынады, яғни:

$$D=max\ /F(x)^{*эмп} - F(x)/\ . \qquad (6)$$

А.Н.Колмогоровтың тұжырымдауынша, егер тәуелсіз тәжірибелер саны "n" шексіздікке қарай өсетін болса, $P(D\sqrt{n} \geq \lambda)$ ықтималдығы мына шекке ұмтылады:

$$P(\lambda) = 1 - \sum_{k=-\infty}^{\infty} (-1)^{\kappa} \cdot \ell^{-2\kappa^2\lambda^2}\ . \qquad (7)$$

Ықтималдық теориясы мен математикалық статистика оқулықтарының "қосымшаларында" (7) өрнекпен есептелген әрбір 0,1 қадам сайын $0 \leq \lambda \leq 2$ мәніне сәйкес, $P(\lambda)$ ықтималдығының шамасы көрініс тапқан кестелер берілген [24, 80, 81].

Егер тәжірибелік үлестірілу функцияның мәндері (жиілік сөздіктегі қатынастық жиіліктер жиынтығы) және теориялық үлестірілу функция мәндері (Вейбулл өрнегімен анықталған қатынастық жиілік жиынтығы) белгілі болса, онда Колмогоров критерийін мынадай реттегі әдіспен қолданғанды жөн санаймыз:

1. Қатынастық жиынтық жиіліктерге сәйкес келетін эмпирикалық және теориялық жиіліктер мәндерінің айырым модульдерінің максимумдық шамасын анықтау:

$$D = \max[\ f_{(x)}^{*эмп.} - f_{(x)}^{*Teop.}\ ].$$

2. Жиілік сөздіктің ең соңғы рет санына (n) қатысты теориялық жиілігі есептелген болса, $\lambda = D(n)$ шамасын анықтау.

3. Белгілі "λ" мәні бойынша, арнайы кесте арқылы $P(\lambda)$ ықтималдық шамасын табу.

4. Егер анықталған $P(\lambda)$ ықтималдық шамасы барынша аз болса, "H_0" болжам қабылданбайды, ал, керісінше, $P(\lambda)$ шамасы айтарлықтай үлкен болса, тәжірибелік пен теориялық үлестірілулер заңдарының үйлесімділігін, яғни H_0 болжамды қабылдауға болады.

Мысал ретінде М.Әуезовтің "Абай жолы" романы кітаптарының (4 кітап, 5 сөздік) мәтінінің сөзтұлға сөздіктерін қарастырайық. Болжам H_0 – роман мәтіндерінен түзілген жиілік сөздіктері негізінде анықталған сөзтұлға жиіліктерінің үлестірілуі Вейбулл функциясы мен берілген тәжірибелік үлестірілу аралықтары үйлеседі. Болжамды тексеру мақсатымен жоғарыда келтірілген әдіс ретімен есептеулер жүргізілу қажет. Арнайы компьютерлік бағдарлама негізінде 5 жиілік сөздіктер деректері бойынша есептеу нәтижелері *3.19-кестеде* орын алды.

3.19-кесте

Параметрлер	Сөздіктер				
	1	2	2	4	5
D	0,0351	0,0336	0,0324	0,0309	0,042
λ	2,001	2,083	1,944	1,944	4,116
$P(\lambda)$	0,001	0,001	0,002	0,002	0,001

Кесте деректерінен байқайтынымыз, роман бойынша түзілген 5 түрлі сөздіктерде де $P(\lambda)$ ықтималдығы нөлге жуық сандар. Сондықтан, тәжірибелік жиіліктер үлестірілуі Вейбулл функциясының үлестірілуімен үйлеспейді, яғни H_0 болжам қабылданбайды. Бұлай тұжырымдаудың негізгі себебі – жиілік сөздік бойының алғашқы зонасы есепке алынды. Сынаққа алынған барлық сөздіктер бойынша айырым модулінің максимум шамасы сөздіктің бастапқы зоналарына сәйкес келеді. Ал егер есептеу процесіне жиілік сөздіктердің алғашқы 50-55 реттік сандарын қоспаған жағдайда $P(\lambda)$ ықтималдығы бірден жоғары көтерілетінін байқаймыз *(3.20-кесте).*

3.20-кесте

Параметрлер	Сөздіктер				
	1	*2*	*3*	*4*	*5*
D	*0,0107*	*0,0103*	*0,0106*	*0,0106*	*0,020*
λ	*0,612*	*0,641*	*0,634*	*0,673*	*1,962*
$P(\lambda)$	*0,864*	*0,864*	*0,864*	*0,711*	*0,502*

$P(\lambda)$ ықтималдығының аз шама қабылдауына сөздіктің соңғы зоналарының әсері бар екендігін ескере отырып, мынадай қорытынды жасауға болады:

Тәжірибелік қатынастық жиынтық жиіліктер үлестірілуінің функциясы ($f_i^{*эмп}$) Вейбуллдың теориялық қатынастық жиынтық жиілігімен $f_i^{*Teop.}$ тек рет саны 50-ден көп және тәжірибелік абсолютті жиіліктің шамасы 5-тен жоғары болған жағдайларда ғана "үйлеседі" (согласуется) екен.

Төртінші тарау
ТІЛДІҢ ЛЕКСИКА-МОРФОЛОГИЯЛЫҚ ҚҰРЫЛЫМЫНА СТАТИСТИКАЛЫҚ ӘДІСТІ ҚОЛДАНУДЫҢ АЛҒЫШАРТТАРЫ

4.1. Тілдік бірліктерді кодтау принципі

Қазіргі қазақ тіліндегі барлық сөздер үлкен үш топқа бөлінеді: *атаушы сөздер, көмекші сөздер, одағай сөздер* [98, 126]. Бұл топтар, әрі қарай лексика-грамматикалық жақтарына сараланып, атаушы сөздер әуелі есімдер және етістіктер деп бөлінеді де, сосын есімдердің өздері іштей атаушы есімдер және үстеуші есімдер болып тағы да жіктеледі. Атаушы есімдер іштей зат есім, сын есім, сан есім, есімдік деген сөз таптарына сараланады да, ал үстеуші есімдер – үстеу сөздер мен еліктеу сөздерге бөлінетіні белгілі. Сөйтіп, жинақтап айтқанда, қазіргі қазақ тіліндегі сөздер 9 топқа бөлініп, олар "сөз табы" деп аталады: зат есім, сын есім, сан есім, есімдік, етістік, үстеу сөз, еліктеу сөз, көмекші (шылау) сөз, одағай.

Сөз табының мазмұны "лексика-грамматикалық" деп аталатын екі компоненттің бірлігінен құралады. Бұл атаудан лексикалық семантикамен бірге грамматикалық семантика деген ұғымды да бірге қабылдау керек. Ал қос сөздің екінші бөлігіндегі "грамматикалық" сөзінің мазмұнына белгілі бір сөзге тән грамматикалық категориялардың және олардың жасалу, түрлену тұлғаларының мағыналары енеді. Яғни бұл жердегі "грамматикалық" ұғымына сөздің жаңадан сөз тудыру, сөз түрлендіру, сөз байланыстыру жүйелерінің мағыналары мен тұлғалары, демек, бүкіл сөздің морфологиялық және синтаксистік белгілері (сыр-сипаттары) түгел енеді [98, 126-134].

Қазақ тілінің қолданбалы саласының, дәлірек айтқанда, статистикалық лингвистиканың алдына қойған мақсаттарының бірі – сөз табының лексикалық (семантикалық), грамматикалық (морфологиялық және синтаксистік) белгілері арқылы, яғни осы үш белгінің негізінде статистикалық зерттеулер жүргізіп, сөздің мағынасын, морфологиялық формасын (тұлғасын) және синтаксистік қызметін сандық деректер арқылы тани білу. Соның нәтижесінде мәтін мазмұнын ашудың формальды жақтарын айқындау.

Сөз таптарының ішкі жіктелген бөліктерінің түрлері тілде біреуі аз не көп қолдануы мүмкін. Статистикалық зерттеу тәжірибесі бойынша олардың қолдану жиілігін анықтау үшін кейбір шараларды алдын ала белгілейтін бағдарламаларды іске қосу қажет.

Мысалы, әртүрлі тілдік стильдердегі сөз таптарының қатынастық сипатының статистикасын білу керек болса, әр сөз табына қатысты шартты белгіні (кодты) қоюдың зерттеушіге ыңғайлы жолдарын ойластыруымыз керек. Себебі, сөз таптарын не басқа тілдік бірліктерді мәтін ішінде түр-тұрпатына қарай формальды әдіспен

ажырату мүмкіндігі әлі де болса зерттеу аясынан тыс қалып жүр. Сондықтан статистикалық санақтарды компьютер арқылы жүргізуде аталған тілдік бірліктерді және олардың ішкі бөліктерін шартты белгілерге сәйкестендіру (кодтау) танымдық рөл үшін аса қажетті. "Кодтау" принципі алдын ала ойластырылған әмбебаптық сипатта болғаны жөн.

Сондықтан қазақ тілінің жоғарыда аталған сөз таптарын ажырату мен олардың статистикасын анықтау мақсатындағы зерттеулерде біздің төменде қарастыратын "кодтау" үлгісін пайдалануға болады:

1) цифрлар арқылы: зат есім – 01, сын есім – 02, сан есім – 03, есімдік – 04, етістік – 05, есімше – 06, көсемше – 07, үстеу – 08, шылау – 09, еліктеуіш сөз – 10, модаль сөздер – 11;

2) әріптер арқылы: зат есім – зт, сын есім – сн, сан есім – са, есімдік – ес, етістік – ет, есімше – еш, көсемше – кш, үстеу – үс, еліктеуіш сөздер – ел, модаль сөздер – мд;

3) цифрлар мен әріптерді араластыра белгілеу арқылы (аралас әдіс): зат есім – з1, сын есім – с1, сан есім – с2, етістік – е1, есімдік – е2, есімше – е3, еліктеуіш сөздер – е4, көсемше – к1, үстеу – ү1, модаль сөздер – м1.

Шартты белгілерді (кодтарды) қоюдың мүмкіндігі мол, дегенмен оларды әркім өз қалауынша белгілегеннен гөрі бірізділікке (стандартқа) ұмтылған дұрыс деп саналады. Әр сөз табының іштей морфологиялық, синтаксистік және семантикалық тармақтану түрлері мен жасалу жолдары да әр зерттеушінің алға қойған мақсатына сай әртүрлі сипатта болуы мүмкін. Сондықтан әрбір сөзге, сөзтұлғаға, сөзқолданысқа, сөзтіркеске немесе мәтіннің одан да үлкен бірліктеріне қойылатын "белгі-код" әртүрлі грамматикалық ақпаратты өзінде сақтай алады. Шартты белгі-кодтарды қабылдау тек бір ғана тілдік бірліктің бірнеше ақпараттарын сәйкестендіретін "ұялар" тізбегі ретінде ұйымдастыруға болады.

Әрине, көбінде, мәтін ішіне мұндай кешенді белгі-код енгізу статистикалық ізденістің бағыт-бағдарына байланысты атқарылады. Қазақ тілін статистикалық әдіспен зерттеу тәжірибесінде тіліміздегі сөз таптарының қолдану сипатын ажырату үшін (жиілігін) мәтінге шартты белгілер енгізудің түрлі мүмкіндіктері жоғарыда көрсетілді.

Енді осындай белгілеу әдісін күрделендіріп, әр сөз табының морфологиялық типтері мен құрылымы жағын ажырату, жасалу жолдарын ескеру, яғни сол тілдік бірлікке қатысты басқа да грамматикалық ақпараттарды барынша қамту және олардың әрбіреуінің статистикасын анықтау мақсатымен, әрі қарайғы баяндауымызда "кодтаудың" кейбір үлгілеріне толығырақ тоқталамыз.

Орыс тілінің морфологиялық сипатына тән сөз таптарын шартты түрде белгілеу (кодтау) ресейлік ғалым Б.Н.Головиннің "Язык и статистика" атты кітабында да сөз болады [30].

Ал морфологиялық ақпарат белгілерін сәйкестендіру бағдарламасын құрудағы біз ұсынып отырған үлгінің ерекшелігі – қазақ тілі сөз таптарының ішкі бөліктерінің құрылымдық сипаты мен жасалу жолдары ескеріліп, әмбебаптық сипатты көздейді.

Бағдарлама жасау ісінің нысаны ретінде – зат есім, етістік, сын есім, сан есім, есімдік сияқты қазақ тілінің негізгі сөз таптары алынды. Аталған сөз таптарына морфологиялық ақпаратты белгі-код үлгісін сәйкестендіру бағдарламасы тиісті кестелер арқылы көрініс тапты.

4.2. Зат есімнің лексика-морфологиялық құрылымына ақпараттық белгі-кодты сәйкестендіру бағдарламасы

Зат есім сөздерге ғана тән морфологиялық ерекшеліктер жоқ емес. Олар өздерінің лексика-семантикалық сипаттарына қарай арнайы жалғаулар арқылы түрленіп және жұрнақтар арқылы тұлғалық өзгеріске ұшырап, сөйлемдегі өзге сөздермен еркін қарым-қатынасқа түсіп отыратыны белгілі. Зат есім сөздердің морфологиялық, синтаксистік және семантикалық белгілерін ұштастыра қарағанда ғана олардың сыр-сипатын толық ашуға болады. Сол белгілер қатарына олардың мәтін ішінде жиі не сирек қолдану сипаты да жатады.

Қазіргі қазақ тілінің грамматикасында зат есім сөздерге ғана тән семантикалық және грамматикалық ерекшеліктері бар мынадай топтар бөліп қарастырылады: адамзат (кімдік) және ғаламзат (нелік) есімдер, жалқы есімдер, көптік мәнді есімдер, эмоциялы-экспрессивтік зат есімдер, көмекші есімдер. Зат есім сөздерді олардың морфологиялық сипаты, яғни құрылымдық, жасалу тәсілі және түрлену жүйесі арқылы да топтап қарастыруға болатыны да белгілі. Міне, осылайша зат есім сөздерді топ-топқа бөліп қарастыру қазақ тілін компьютерлік және статистикалық лингвистика әдістерімен зерттеуде аса маңызды деуге болады.

Кітаптың бірінші тарауында айтылғандай қазақ тіліндегі мәтіндерде ең жиі қолданыс табатын сөз таптары *зат есім* мен *етістік*. Олай болатын болса, олардың мәтін ішінде морфологиялық, синтаксистік жолмен түрленіп қолдануының да статистикасын танып білу "Қолданбалы лингвистика" пәнінің мақсаттарының бірі. Біздің баяндауымыздағы негізгі ұстаным – мәтін ішінде кездесетін *сөз*, *сөзтұлға* не *тұрақты сөзтіркеске*, яғни мәтін бірлігі деп аталатын – *сөзқолданысқа* сәйкес келетін грамматикалық ақпараттар *орны ("ұясы")* мен сол ақпаратты танытатын шартты *белгі-кодты* тұрақтандыру. Қандай да болмасын тілдік бірліктің, мысалы, ең алдымен қарастырайын деп отырған – *зат есім* сөздердің әрбіреуіне сәйкестендірілетін шартты параметрлерді ("ұя" орны мен белгі-код) алдын ала жан-жақты ойластырып алғанды жөн санаймыз.

Айталық, *зат есім* сөздерге сәйкестендіретін (тиісті) "ұя" саны немесе топтама саны – сегізге тең делік, яғни n=1, 2, 3, 4, 5, 6, 7, 8. Ал әрбір "ұяда" жазылатын шартты *белгі-код* саны (топтама саны) мен "код мәні" *зат есім* сөздердің сол "ұяда" қарастырылатын сипатына қарай өзгеріп отыруы мүмкін.

Егер осы айтылғандарды қамтитын мәліметтер кестесін *"бағдарлама"* деп атасақ, *зат есім* сөз табына қатысты қысқаша топтама *4.1-кестедегі "бағдарлама"* түрінде көрініс табады.

Топтамадағы мәліметтердің кейбіреулеріне түсініктеме берейік.

Аталған кесте үш "бағанадан" тұрады:

1) *шартты белгілер орны* (топтама саны);

2) *зат есім сөздердің лексика-морфологиялық сипаты*;

3) *шартты белгі-код.*

Әрбір сөз табына қатысты "ұялар" топтамасының ең алғашқысы сол сөз табының қысқарған атымен белгіленеді (кодталады) және *1-ші орынға* ие болады деп ұйғару

керек. Сонымен, "ұялар" тізбегінің ең алғашқысы (бірінші бағана) "1" санына тең (бірінші топтама), осы жатық жолдағы екінші бағана сөз табының сипаты жайлы мәлімет, яғни – *зат есім* атауы және үшінші бағана – мәліметтің "белгі-код" мәні – "ЗТ". Сол сияқты шартты белгілер орнының "2" санында, яғни екінші топтамасында *зат есім* сөздердің екі түрлі сипаттамасы берілген. Олардың біріншісі – *адамзат есімдері (кім?)* және екіншісі – *ғаламзат есімдері (не?)*. Шартты белгі орнында, яғни 2-ші орында қарастырылатын кезекті *зат есім* сөздің қай сұраққа *(кім? не?)* жауап беруіне қарай үшінші бағанадағы оған сәйкес келетін белгі-кодтың біреуі ғана таңдалып алынады. *4.1-кестеде* зат есім сөздерге қатысты осы тәріздес 8 топтама қарастырылды.

Сол топтамалардың енді 7-ші орынға қатыстысын сөз етсек, жетінші топтамада *зат есім* сөздің алты түрлі сипаты аталған: *1. Түбір морфема зат есім; 2. Туынды зат есім (түбір+жұрнақ); 3. Күрделі зат есім: а) біріккен сөз; ә) қосарланған сөз; б) құрама сөз; в) қысқарған сөз.*

Үшінші бағанада олардың белгі-код мәндері үш орынды сандар: 070, 071, 072, 073, 074, 075 сандары арқылы берілген (*4.1-кестені* қараңыз).

Жоғарыда 2-ші орынға байланысты айтылған сияқты, бұл жолы да зерттеуге алынған кезекті сөзді (сөзтұлғаны) шартты түрде "кодтау" кезінде 7-ші орынға сәйкес келетін (070 пен 075 аралығы) алты белгі-код мәндерінің ішінен сөзтұлға сипатына тән келетін біреуі ғана алынады. Кестеде келтірілген басқа топтама орындарына да осылайша түсініктеме беруге болады.

Енді мысал ретінде кейбір зат есім сөздерді *4.1-кесте* деректері бойынша шартты белгі-код түрінде жазып көрейік. Мысалы, мына сөйлем ішіндегі зат есім сөздерді бөліп алып, белгі-код арқылы жазу керек: *Асан мектепке барарда кітап, қалам, дәптер және басқа да керек-жарақтарын аса ұқыптылықпен сөмкесіне салды.* Бұл сөйлемдегі зат есім сөзтұлғалар мыналар: *Асан; мектепке; кітап; қалам; дәптер; керек-жарақтарын; сөмкесіне.*

Шартты белгі-кодтарды осы сөздерге сәйкестендірейік. Яғни әрбір аталған *зат есім* сөздердің сегіз орындық "ұяларына" тиісті *белгі-код* мәнін қойып шығу керек деген сөз. Олардың барлығының да бірінші орындарында "ЗТ/" белгі-код тұратыны айғақ.

Енді *Асан* сөзін бөлек алып қарастырсақ, 2-ші орында – "020" (адам аты болғандықтан), үшіншіде – "030", төртіншіде – "040", бесіншіде – "050", алтыншыда – "060", жетіншіде – "070", сегізіншіде – "080". Сонымен, сөйлемдегі *Асан* сөзі мен басқа да зат есім сөздердің "кодтық" жазылуы төмендегідей болады:

Асан:	ЗТ/	020	030	040	050	060	070	080
мектепке:	ЗТ/	021	031	040	050	060	070	083
кітап:	ЗТ/	021	031	040	050	060	070	080
қалам:	ЗТ/	021	031	040	050	060	070	080
дәптер:	ЗТ/	021	031	040	050	060	070	080
керек-жарақтарын:	ЗТ/	021	031	042	050	060	073	084
сөмкесіне:	ЗТ/	021	031	040	050	060	070	083

4.1-кесте

Зат есім сөздерге лексика-морфологиялық ақпаратты сәйкестендіру бағдарламасы

Белгі-код орны	Зат есім сөздердің лексика-морфологиялық сипаты	Белгі-код
1	Зат есім сөз (сөзтұлға)	ЗТ/
2	Адамзат есімдері *(кім?)*	020
	Ғаламзат есімдері *(не?)*	021
3	Жалқы есім: а) кісі аттары (ономастика); ә) географиялық атаулар (топонимика)	030
	Жалпы есім	031
	Аралас мағыналы *(ай, күн, жер)*	032
4	Көптік мәнді емес (жалғаусыз қарастырғанда)	040
	Табиғи қос мәнді *(аяқ, құлақ, етік)*	041
	Көптік мәнді (жалғаусыз)	042
5	Реңсіз зат есім	050
	Эмоциялы-экспрессивтік реңді зат есім *(Әкей, Сәулеш)*	051
6	Көмекші зат есімге жатпайтын зат есім	060
	Көмекші зат есім *(алды, арты, қасы)*	061
7	Түбір морфема зат есім	070
	Туынды зат есім (түбір+жұрнақ)	071
	Күрделі зат есім: а) біріккен сөз	072
	ә) қосарланған сөз	073
	б) құрама сөз	074
	в) қысқарған сөз	075
8	Зат есімнің түрленуі:	
	Зат есімнің жалғаусыз түрі *(кім? не?)*	080
	Көптік жалғаулармен: *-лар, -лер,-дар,-дер,-тар,-тер*	081
	Тәуелдік жалғаулар арқылы: *-нікі, -дікі, -тікі, -м, -ым, -ім; -ң, -ың, -ің, -ңыз, -ңіз, -ыңыз, - іңіз; -сы, -сі, -ы, -і*	082
	Септік жалғаулар арқылы: ілік *(кімнің? ненің?);* барыс *(кімге? неге? қайда?);* табыс *(кімді? нені?);* жатыс *(кімде? неде?);* шығыс *(кімнен? неден?);* көмектес *(кіммен? немен?).*	083
	Жіктік жалғау арқылы: жекеше*: -мын, -мін, -сың, -сің, -сыз, -сіз;* көпше*: -мыз, -міз.*	084

Көрсетілген жолмен мына зат есімдерге қатысты сөзтұлғаларды белгі-кодпен сәйкестендірейік: *ғалым, Асандар, құлағым, Алматыға, жазушы, саңырау құлақтан, жаным-ай, жүрісіме.*

ғалым	ЗТ/	020	032	040	050	060	070	080
Асандар	ЗТ/	020	030	040	050	060	070	081
құлағым	ЗТ/	021	032	041	050	060	070	082
Алматыға	ЗТ/	021	030	040	050	060	070	083
жазушы	ЗТ/	020	032	040	050	060	071	080
саңырау құлақтан	ЗТ/	021	032	041	050	060	074	083
жаным-ай	ЗТ/	021	032	040	051	060	070	082
жүрісіме	ЗТ/	021	032	042	050	060	071	083

Белгілі бір мәтін сөзтұлғаларын белгі-код түріне ауыстырып барып, жиілік сөздіктер түзуге болады немесе әріп күйіндегі мәтіннің өзінен түзілген әліпби-жиілік, жиілік не кері-әліпби сөздіктердегі сөзтұлғаларды (бірліктерді) белгі-код түріне ауыстырып жазуға да болады. Егер ондай сөздіктерді тиісті компьютерлік бағдарлама бойынша "қысылған" түрге келтіріп жазатын болсақ, зат есім (не басқа сөз табы) сөзтұлғалардың "сипаттық жиілік сөздігін" алуға болады. Бұндай жиілік сөздік қазақ тілінің морфологиялық құрылымын зерттеу нысаны ретінде және әрі қарайғы мәтін лингвистикасына қажетті маңызды материал болатыны сөзсіз.

4.3. Етістіктің лексика-морфологиялық құрылымына ақпараттық белгі-кодты сәйкестендіру бағдарламасы

Етістік сөздер құрамының басқа сөз таптарына қарағанда күрделілігі оның аса өрісті лексика-семантикалық сипатымен, өте бай лексика-грамматикалық формаларымен және кең синтаксистік қызметімен тығыз байланысты. Етістіктің лексика-грамматикалық тұлғаларының бай болу себебі – олар әрекеттің болу мезгілін, жүзеге асу кезеңін, өту сипатын, яғни әрекеттің бағыты, қарқыны, тынуы тәрізді жайларды түгел қамтиды. Осымен қатар басқа сөз табына қатысты сөздерден етістік тудыратын синтетикалық, аналитикалық тәсілдер жүйелерін қоса есептесек, етістіктің тұлғалану байлығына жететін сөз табы жоқ деуге болады. Осы айтылғандар оның өзіне ғана тән лексика-семантикалық, лексика-грамматикалық және грамматикалық категорияларынан айқын көрініп жатады. Етістіктің осындай әрқилы категорияларының мағыналары да, синтаксистік қызметтері де өзге сөз таптарына тән сөздермен қарым-қатынасқа түскенде айқындала түседі.

Етістіктің лексикалық жүйесіндегі әрбір сөздің өзіне ғана тән лексикалық мағыналары бөлек болғанымен, семантикалық ерекшеліктеріне қарай бір сөз тобына телінеді. Етістіктің грамматикалық тұлғалары мен қызметтерін дұрыс айқындау үшін, оларды түбір формалардың, туынды синтетикалық және аналитикалық формалардың семантикалық құрылымына орай жіктеп, топтап барып сыр-сипатын ашу, олардың тұлғалық және мазмұндық арақатынастарын айқындауға әкеліп соғады. Себебі грамматикалық семантика сөздің құрамына көп байланысты болып келеді.

Сол себептен етістік формаларының морфологиялық құрылымының сыр-сипатын ашу мәнмәтіндік тұрғыда немесе мәтіндік лингвистика саласының зерттеу аясына жақындайды. Ал бұл тұрғыдан зерттеуді ұйымдастыру дегеніміз статистикалық лингвистика әдістерін молынан қолдану деген сөз. Туынды етістіктердің семантикалық құрылымының аса күрделі және түрлену сипатының бай болу себептерінің бірі – олардың сирек не жиі қолдануында деп түсіну қажет [70].

Сондықтан туынды етістіктердің морфологиялық құрылымын жеке бір жүйе ретінде қарастырып, әрбір өзіне тән сипатына қарай топтап, олардың ықтималды-статистикалық заңдылықтарын ашу үшін арнайы белгі-код жүйесін енгізіп зерттеу қажет деп санаймыз. Себебі, жоғарыда, зат есім сөздерге қатысты айтылғандай, қазақ тілінің тілдік бірліктерін, сөз таптарын формальды белгілері арқылы танып білу және соның нәтижесінде автоматты түрде жазба не сөйлеу мәтінінен оларды бөліп алу мәселесі өз шешімін таппай отыр. "Қолдан" енгізген шартты белгі-код негізінде етістіктің морфологиялық құрылымына статистикалық зерттеулер жүргізу көптеген жайттардың басын ашатыны белгілі. Солардың ішінде мәтіндік бірліктерді автоматты түрде айыра білу мен бөліп алу мәселесі де бар. Сонымен бірге етістіктің лексика-грамматикалық құрылымының дәстүрлі әдіспен зерттеуде байқала бермейтін тұстарының да басы ашылуы (анықталуы) мүмкін. Қазіргі қазақ тіліндегі етістіктерді (түбір және туынды) осы тұрғыда қарастыра отырып, біз төменде олардың іштей бір-біріне мағына жағынан жақындықтарына, өзара функция жағынан орайластықтарына және лексика-морфологиялық құрылымына қарай топ-топқа бөліп, етістік сөздердің сипаты жайлы ақпараттарды алдын ала ойластырылған шартты белгі-кодтарға сәйкестендіру бағдарламасын ұсынамыз.

Мысалы, мағыналық тұрғыдан алып, етістік сөздерді іштей мына топтарға бөлуге болады [98, 223-226]: **амал-әрекет, қимыл-қозғалыс, қалып-сапа, ойлау-сөйлеу, өсу-өну етістіктері** және олар *4.2-кестеде* көрсетілгендей жалғасын табады.

Әдетте, туынды сөздердің бір жүйе бойынша қалыптасқан белгілі морфологиялық құрылымы болатыны белгілі. Сол құрылымның бірінші компоненті дербес мағыналы сөз болады да, ал екінші компоненті – бірінші компонентті белгілі бір сөз табына айналдыратын жұрнақ болады *(міс+те, кел+тір, сабын+да)*. Туынды етістіктер, жасалатын негіздеріне қарай, есімдерден және етістіктерден жасалған етістіктер деген екі салаға бөлінеді. Олар есім негізді етістіктер, етістік негізді етістіктер деп аталып жүр. Біз осылардың тек алғашқысын ғана, яғни есім негізді етістіктердің жасалу жолдарын жекелеп 16 топқа бөліп қарастырып, оларды шартты белгі-код арқылы таңбалау үлгісін *4.2-кесте* арқылы ұсындық.

Зат есім сөздерге лексика-морфологиялық ақпаратты сәйкестендіру бағдарламасы төріздес, *4.2-кестедегі* бағдарлама да үш бағанадан тұрады:

1) белгі-код орны; 2) есім негізді туынды етістік сөздердің сипаты; 3) белгі-код таңбасы.

Бірінші бағананың белгі-код орны "1" (немесе №1-ші топтама), ал екінші бағана бойынша оның сипаттамасы – "етістік сөз" деп аталады. Егер мәтін бойынан зерттелетін кезекті сөз – *етістікке* жататын болса, онда үшінші бағанадан сол етістік сөздің *белгі-коды* – "ЕТ" болады. Екінші орынға немесе екінші топтаманың екінші бағанасында етістіктердің өздерін іштей мағыналық және өзара функционалдық жағынан жақындастыратын 11 топтың әрбіреуіне сипаттама берілген. Ал үшінші бағанада осы сипаттамаларға сәйкестендірілген белгі-код таңбасын айыра білуге болады. Мысалы, белгі-кодтың "2-ші" орнының (2-ші топтамасының) үшінші бөлігі *қалып-сапа етістіктері* деп аталады, ал оған сәйкес келетін үшінші бағанада сол сипаттаманың үш орынды белгі-коды – "102" тұрады. Сол сияқты, туынды етістіктердің жұрнақтар арқылы есім сөздерден жасалу жолынан да мысал келтірейік. Туынды етістіктердің ***-ла (-ле, -да, -де, -та, -те)*** жұрнағы арқылы жасалу жолы ішкі 5 бөлікпен топталған атаулар сипатымен берілген:

1) дене мүшелері (*аяқта, жұдырықта, өкшеле*);

2) еңбек құралдары (*арала, балтала, арқанда);*

3) іс-әрекетті жүзеге асыруға объект болатын зат атаулары (*жүнде, майла, тұзда, алтында);*

4) мекен я орын, өлшеу, дыбыс т.т. (*мекенде, өрле; аршында, метрле; мыңқылда, шыңқылда);*

5) сын есім, сан есім, үстеу, еліктеуіш т.т. сөздерге жалғанған жағдайлар (*ақта, қарала, екеуле, аһла, уһле).*

Осы 5 түрлі топтарға 5 үш орынды сандар (белгі-код) сәйкестендірілген: 113, 114, 115, 116, 117. Осы тәріздес, басқа топтама орындарында да өзіне тән сипаттамалар және соған сәйкес шартты белгі-кодтар *4.2-кестеде* көрініс тапты.

Енді осындай белгі-кодтар жиынын іс жүзінде қалай пайдаланамыз деген сұрақ туындауы мүмкін. Бұл жайында жоғарыда зат есім сөздерге қатысты да қысқаша айтылған болатын. Етістік сөздерге қатысты *4.2-кестені* пайдалануда да айтарлықтай айырмашылық жоқ деуге болады. Осы тұста айта кетерлік жайт, топтама орындар санын (белгі-код орнын) тағайындау (белгілеу), біріншіден, зерттеу мақсатына сай, екіншіден, әр сөз табының лексика-морфологиялық, семантикалық ерекшеліктеріне байланысты деуге болады. Мысалы, зат есім сөздерге қатысты *4.1-кестеде* сегіз орын жеткілікті деп алсақ, етістік сөздерге қатысты *4.2-кестеде* ондай орындар саны төртеу-ақ. Ал басқа сөз табына қатысты немесе зерттеу мақсатына сай, топтау орын саны басқаша болуы да мүмкін. Сол сияқты, *4.1* және *4.2-кестелердегі* деректерден байқайтынымыз әр топтаманың ішкі бөлік саны да әртүрлі болады екен. Енді осы аталған кестелерді пайдалану жайына қайта оралайық. Алғашында, зерттеуге алынған мәтін бөлігіндегі сөзқолданыстарды сөз табына ажырататын шартты белгілер қойылғаннан кейін, олардан компьютер арқылы қажетті жиілік сөздіктер алынуы керек. Аталған жиілік сөздіктегі етістік сөзтұлғаларды (не басқа сөз табына қатысты сөздерді) бөліп алып, оларға *4.2-кесте* бойынша белгі-кодтарды сәйкестендіру қажет болады. Ескерту ретінде айта кететін жайт, зерттеуді жиілік сөздіксіз-ақ бірден мәтінмен де жүргізуге болады, бірақ бұл тәсілде қайталанатын мәтін бірліктері көп уақыт алады. Қалай болғанда да зерттеу жұмысы тізбек құрайтын сөзқолданыс жүйесінен етістік сөзтұлғаларды (немесе басқа сөз табы) кезек-кезегімен бөліп алып қарастыруды қажет етеді.

4.2-кесте

Етістік сөздерге лексика-морфологиялық ақпаратты сәйкестендіру бағдарламасы

Белгі-код орны	Етістік сөздердің лексика-морфологиялық сипаты	Белгі-код
1	2	3
1	**Етістік сөз**	ЕТ/
2	**Мағыналық және функционалдық жағынан топтастыру:**	
	1) Амал-әрекет етістіктері Мысалы: *босат, көтер, күрес, қи, сыз, өлше* т.т.	100
	2) Қимыл-қозғалыс етістіктері Мысалы: *ауна, аудар, домала, секір, бүкірей* т.т.	101
	3) Қалып-сапа етістіктері. Мысалы: *жат, жантай, тұр, отыр, ұлғай* т.т.	102
	4) Ойлау-сөйлеу етістіктері. Мысалы: *айт, сөйле, де, ескер, жатта, ұмытпа* т.т.	103

4.2-кестенің жалғасы

1	2	3
	5) Өсу-өну етістіктері. Мысалы: *балала, жапырақта, гүлде, өс, қозыла* т.т.	104
	6) Бағыт-бағдар етістіктері. Мысалы: *бар, кел, қайт, әпер, әкет, қаш, түс, көтер* т.т.	105
	7) Көңіл-күй етістіктері. Мысалы: *жыла, қайғыр, өкін, күл, қуан, алақайла* т.т.	106
	8) Бейнелеу-еліктеу етістіктері. Мысалы: *жарқыра, күркіре, дүркіре, тарсылда* т.т.	107
	9) Дыбыс-сес етістіктері	108
	10) Көру-есту етістіктері.	109
2	11) Мінез-құлық етістіктері	110
	12) Етістіктің тізімге енбеген түрлері	111
3	**Түбір етістік** **Анықтама:** Етістік түбірі неше қилы формалар (есімше, көсемше, рай, етіс, шақ т.б.) тудыратын қосымшалардың барлығын алып тастағанда сақталатын бөлігі. Мысалы: *аз, ал, айт, алда, ат, ас* т.б.	112
	Есім негізді сөздерден жасалатын туынды етістіктер. Жұрнақтар арқылы жасалу жолдары: 1. ***-ла (-ле, -да, -де, -та, -те)*** жұрнағы арқылы жасалған туынды етістіктер: а) дене мүшелері атауларына жалғанады. Мысалы: *аяқта, жұдырықта, өкшеле.*	113
	ә) Еңбек құралдары атауларына жалғанады. Мысалы: *арала, балтала, арқанда.*	114
	б) Іс-әрекетті жүзеге асыруға объект болатын заттарға жалғанады. Мысалы: *жүнде, майла, тұзда, алтында.*	115
	в) Мекен я орын, өлшеу, дыбыс, қозғалыс, көрініс, төл және әлеуметтік, саяси, мәдени, тұрмыс, салт, табиғат т.т. атауларына жалғанады. Мысалы: *мекенде, өрле, төменде; аршында, метрле, тоннала; мыңқылда, шыңқылда; жарқылда, бұрқылда, ирелеңде; ботала, қозда; тәрбиеле, еркеле, жазала, қаумала, үймеле* т.т.	116
	г) Сын есім, сан есім, үстеу, еліктеуіш, одағай сөздерге жалғанады. Мысалы: *ақта, қарала, жаманда; онда, жүзде, екеуле; төменде, кейінде, ілгеріле; жымыңда, күлімде; аһла, уһле, ойбайла* т.т.	117
	2. ***-лан (-лен, -дан, -ден, -тан, -тен)*** жұрнағы арқылы жасалады. Мысалы: *ашулан, арлан, борышлан, иелен ...*	118
	3. ***-лас (лес, -дас, -дес, -тас, -тес)*** жұрнағы арқылы жасалған. Мысалы: *бірлес, көмектес, бәстес, достас, кезектес, сабақтас* т.б.	119

4.2-кестенің жалғасы

1	2	3
	4. ***-лат (-лет, -дат, -дет)*** жұрнағы арқылы жасалған. Мысалы: *дауылдат, тездет, тұнделет* т.б.	120
	5. ***-а (-е)*** жұрнағы арқылы. Мысалы: *аша, деме, жаса, өрте, тіле, сына, міне* т.б.	121
	6. ***-ай (-ей, -й)*** жұрнағы арқылы. Мысалы: *күшей, мұңай, қартай, кеңей, көбей, молай* т.б.	122
3	7. ***-қар (-ғар, -кер, -гер)*** өнімсіз жұрнағы арқылы. Мысалы: *басқар, ескер, аңғар, теңгер, ақыр, қақыр, жазғыр, ысқыр, кекір* т.б.	123
	8. ***-ар (-ер, -р)*** көне жұрнағы арқылы. Мысалы: *жаңар, тазар, ескір, өзгер* т.б.	124
	9. ***-ал (-ел, -ал, -іл, -л)*** қосымшасы арқылы. Мысалы: *жоғал, оңал, тарыл, тіріл, теңел* т.б.	125
	10. ***-ық (-ік)*** көне жұрнағы арқылы. Мысалы: *ашық, бірік, зарық, дәнік, өшік* т.б.	126
	11. ***-сы (-сі)*** және ***-ымсы (-імсі)*** көне жұрнақтары арқылы. Мысалы: *батырсы, босаңсы, көлгірсі, үлкенсі, апамсы* т.б.	127
	12. ***-сын (-сін)*** жұрнағы арқылы. Мысалы: *адамсын, білгішсін, көпсін, жамансын* т.б.	128
	13. ***-сыра (-сіре)*** жұрнағы арқылы. Мысалы: *айсыра, өлсіре, қансыра, жетімсіре* т.б.	129
	14. ***-ыра (-іре)*** жұрнағы арқылы. Мысалы: *барқыра, бұрқыра, дүркіре, күркіре* т.б.	130
3	15. ***-ырай, (-ірей)*** жұрнағы арқылы. Мысалы: *бақырай, шақырай, кішірей, шікірей,* т.б.	131
	16. ***-ы, -і; -шы, -ші; -ан, -ен, -ын, -ін, -н; -ырқа, -ірке, -ырқан, -іркен; -ына, -іне; -қа, -ке, -ға, -ге*** т.б. көне жұрнақтар арқылы. Мысалы: *байы, жасы, желпі, кеңі; аунақшы, дөңбекші; тасырқа, мүсірке, ашырқан, шіміркен; есіне, қатына, пысына; бүрке, шіске, қозға; басын, жирен, оян, үйрен* т.б.	132
	Етістік негізді сөздерден жасалатын туынды етістіктер	133
4	Күрделі етістіктер	140
	Жалаң етістік	141

Белгіленген бір орынға сол топтамадағы белгі-кодтар ішінен біреуі ғана таңдалып алынатынын үнемі есте ұстау қажет.

Мысал ретінде мына сөйлемде кездесетін етістік сөзтұлғаларды бөліп алып қарастырайық: *"Кешке әрі тоңып, әрі шаршап аһлап, уһлеп отырған кемпір өзін өлтіріп кете жаздаған кім екенін есіне алды"* (С.Көбеев).

Етістік сөзтұлғаларға жататындар: *тоңып, шаршап, аһлап, уһлеп, отырған, өлтіріп кете жаздаған, алды.*

Енді *4.2-кестеде* көрсетілген деректер бойынша етістіктерге белгі-код мәндерін сәйкестендірейік:

тоңып	ЕТ/	106	133	141
шаршап	ЕТ/	106	133	141
аһлап	ЕТ/	106	133	141
үһлеп	ЕТ/	106	133	141
отырған	ЕТ/	102	133	141
өлтіріп	ЕТ/	100	133	141
кете жаздаған	ЕТ/	105	140	140
алды	ЕТ/	111	133	141

Сонымен, әрқилы етістік сөздер өздерінің жасалу және мағыналық ерекшеліктеріне қарай неше түрлі белгі-код мәндерін қабылдай алады.

4.4. Сын есімнің лексика-морфологиялық құрылымына ақпараттық белгі-кодты сәйкестендіру бағдарламасы

Қазіргі қазақ тілінің морфология саласында сын есім деп аталатын сөз табына мынадай анықтама берілген: *"Заттың сапасын, сипатын, қасиетін, көлемін, салмағын, түсін (түр-реңін) және басқа да сыр-сипаттарын білдіретін лексика-грамматикалық сөз табы сын есім деп аталады"* [98, 166-172]. Морфологиялық сипаттары жағынан сын есімнің өзге есімдерге қарағанда өзіне ғана тән ерекшеліктері болатыны мәлім. Мысалы, зат есіммен, үстеумен және т.б. сөз таптарымен әрі ортақ, әрі тұлғалас сөз тудыратын жұрнақтарымен қатар тек туынды сын есім ғана жасайтын арнаулы жұрнақтары және сын есімнен сын есім тудыратын шырай жұрнақтары да бар. Бұл аталғандар да сын есімнің морфологиялық формалары болып есептеледі.

Сын есімдердің тағы да бір өзіндік қасиеті – заттардың алуан түрлі сыр-сипаттары мен белгілерін тікелей білдірумен қатар басқа заттардың қатынастары арқылы да сондай сипаттарды білдіре алады. Осыған байланысты, сын есімдер семантикалық мағыналары мен грамматикалық ерекшеліктеріне қарай, *сапалық (негізгі)* және *қатыстық (туынды)* сын есім деп аталатын екі салаға бөлінетіні де белгілі.

Негізгі немесе сапалық сын есімдер жайлы айтатынымыз, олар, әдетте, ешбір қосымшасыз тұрып-ақ, заттың әрқилы сыр-сипатын білдіретін түбір сөздерден тұрады. Бірақ, егер кейбір түбір сөз деп жүрген сөздерді олардың тарихи даму тұрғысынан қарастырсақ, олар туынды сөз болып та шығуы мүмкін. Сондықтан, құрылымы жағынан негізгі сын есім деп есептелетін сөздердің саны да, сапасы да ұдайы өзгеріп отырады. Ал сын есімдерді морфемалық құрылымына қарай, негізгі және туынды деп жіктеу шартты екендігін де есте ұстаған жөн.

Сонымен кез келген сөйлем ішінде (мәтінде) кездесетін сын есім сөздер, біріншіден, сапалық сын есімге жатуы мүмкін *(ақ, қара, сары, көк, сұр, биік, үлкен, аласа, жылы, суық, ірі, кіші* т.т.), екіншіден, морфологиялық (синтетикалық) тәсіл бойынша, яғни тиісті жұрнақтар арқылы жасалған туынды сын есімге жатуы мүмкін екен.

Үшіншіден, қарастырылатын кезекті сын есімге қатысты сөз синтаксистік (аналитикалық) тәсіл, яғни жалаң сын есімдердің бір-бірімен тіркесуі арқылы *(ақ*

сары, қызыл сары, ақ шұбар, ақ көйлекті, көп балалы, үлкен-кішілі т.т.) және морфологиялық-синтаксистік (семантикалық) тәсіл бойынша жасалатын туынды сын есімдер де болуы мүмкін.

Соңғы аталған тәсіл бойынша туатын сын есімдер қолданылуы жағынан алғанда тіпті де өнімсіз деуге болады. Сондықтан сын есімдердің жасалу жолдарының ең негізгісі деп, ең алдымен морфологиялықты немесе синтетикалықты, ал одан кейін барып, аналитикалықты, яғни синтаксистік тәсілді айтуға болады.

Осы себептен, біз сын есімнің лексика-морфологиялық құрылымына ақпараттық белгі-кодты сәйкестендіру бағдарламасын құрастыруда тек қана сын есімнің синтетикалық тәсілмен жасалу жолына ғана, дәлірек айтқанда, есімдерден сын есім тудыратын өнімді жұрнақтар арқылы жасалу жолын негіз етпекпіз.

4.3-кесте

Сын есімнің лексика-морфологиялық құрылымына ақпараттық белгі-кодты сәйкестендіру бағдарламасы

Белгі-код орны	Сын есім сөздердің лексика-морфологиялық сипаты	Белгі-код
1	2	3
1	**Егер сөз (сөзтұлға) морфемалық құрамына қарай сапалық сын есім (негізгі сын есім) болса:**	СН1
2	1) заттың не құбылыстың түрі мен түсін анықтаса. Мыс.: *ақ, қара, қызыл, жасыл* т.б.	201
	2) заттың не құбылыстың сыры мен сапасын анықтаса. Мыс.: *жақсы, жаман, тәуір, нашар* т.б.	202
	3) заттың көлемі мен аумағын, ұзындығы мен салмағын анықтаса. Мыс.: *үлкен, кіші, ұзын, ауыр, қысқа, жеңіл* т.б.	203
	4) заттың дәмі мен иісін білдірсе. Мыс.: *ащы, тәтті, күлімсі* т.б.	204
	5) заттың не құбылыстың басқа да қасиет-белгілерін білдірсе.	205
1	**Егер кезекті сөз (сөзтұлға) морфемалық құрамына қарай қатыстық сын есімге жатса, яғни заттың белгісін басқа бір заттың қатысы арқылы білдірсе және ондай туынды сын есімдер өнімді (өнімсіз) жұрнақтар арқылы жасалса.**	СН2
2	Есімдерден туынды сын есім тудыратын өнімді жұрнақтар арқылы жасалып, заттың сыртқы түрі мен түсіне, кескіні мен келбетіне, сыры мен сынына, мекен мен мезгілге және т.б. белгілеріне қатысты сындық ұғым білдірсе. 1. ***-қы, -кі; -ғы, -гі*** жұрнақтары арқылы кейбір зат есімдерден, есімдіктерден, үстеулерден, сондай-ақ, жатыс және шығыс септіктегі сөздерден жасалатын туынды сын есімдер мынадай мағынада жұмсалса:	

4.3-кестенің жалғасы

1	2	3
	а) егер мекендік ұғым білдіретін кейбір зат есімдерге, үстеулерге, сондай-ақ, жатыс (кейде шығыс) септік формаларындағы есімдерге жалғанса. Мыс.: *ауызғы, төргі, түпкі, ішкі, төменгі, соңғы, санаулы, буктеулі* т.б.	210
2	ә) егер мезгілдік ұғым білдіретін кейбір зат есімдер мен есімдіктерге, мезгілдік үстеулерге жалғанса. Мыс.: *кешкі, түскі, күзгі, жазғы, көктемгі, түнгі, күндізгі* т.б.	211
	2. ***-лы, -лі, -ды, -ді, -ты, -ті*** қосымшасы арқылы мынадай туынды сөздер жасалады:	
	а) белгілі бір заттың (я құбылыстың) бар екендігін білдіру үшін зат есімдерге жалғанса. Мыс.: *арлы, сулы, әсерлі, гүлді, икемді, байыпты, инабатты, балалы* т.б.	212
	ә) қосарланған зат есімнен, сын есімнен, сан есімнен, үстеуден күрделі сын есімдер осы қосымшалар арқылы жасалса. Мыс.: *ағалы-інілі, ойлы-қырлы, таулы-тасты, өзенді-сулы, үлкенді-кішілі, бұрынды-соңды* т.б.	213
	б) екі-үш сөзден құралып, суреттеме атаулар қызметіндегі күрделі сын есімдер жасалса. Мыс.: *ақ басты, қаз мойынды, ай қабақты, теке сақалды* т.б.	214
	3. ***-сыз (-сіз)*** жұрнағы есім сөздерге жалғанып, болымсыздық мағынадағы туынды сын есімдер жасалса. Мыс.: *баласыз, көліксіз, білімсіз, сенсіз, бізсіз, мұңсыз* т. б.	215
	4. ***-шыл (-шіл)*** қосымшасы зат есім, есімдік, модаль сөздерге (әр тарап) жалғанып, бейімділікті, икемділікті, құмарлықты білдіретін қатыстық сын есімдер жасалса. Мыс.: *ұйқышыл, ұйымшыл, өзімшіл, турашыл, ойшыл* т.б.	216
	5. ***-дай, (-дей, -тай, -тей)*** жұрнағы жалғанып, сөздерді салыстыру, ұқсату мәнді туынды сын есімдер жасалса. Мыс.: *аттай, әкедей, мендей, сендей, жүздей, өлердей* т.б.	217
	6. ***-лық, -лік, -дық, -дік, -тық, -тік*** жұрнақтары арқылы мына мағынадағы туынды сын есімдер жасалса:	
	а) зат есімдерге жалғанып, сөздердің нақтылылық мағына қасиеттерін білдіретін туынды сын есімдер жасалса.	218

4.3-кестенің жалғасы

1	2	3
	Мыс.: *орталық, қоғамдық, қалалық, азаматтық, жолдастық* т.б.	
2	ә) мезгіл атаулары мен әрқилы бұйым атауларына жалғанып, мезгіл, өлшеу мөлшерімен байланысты туынды сын есімдер жасалса. Мыс.: *айлық, жылдық, апталық, тәуліктік, көйлектік, пальтолық, қайнатымдық, екі-үш асымдық* т.б.	219
	б) есімдіктерге жалғанып, олардан белгілі бір жаққа қатыстықты білдіретін туынды есімдер жасалса. Мыс.: *өздік, мендік, сендік, қандайлық, қаншалық* т.б.	220
	7. ***-лас, -лес, -дас, -дес, -тас, -тес*** жұрнақтары арқылы мына мағынадағы туынды сын есімдер жасалса:	221
	а) адам мінезіне байланысты зат есімге жалғанып, ондай сипаттың басқа адамдарға да қатысты екендігін білдірсе. Мыс.: *ниеттес, сырлас, тілеулес, пікірлес* т.б.	
	ә) адамзат, жан-жануарлар тегімен байланысты ұғымдағы зат есімдермен жалғанып, тұқым-туыстас екенін білдіретін сын есім жасалса. Мыс.: *туыстас, аталас, бауырлас, қарындас, тумалас* т.б.	222
	б) мекен ұғымдарына жалғанып, олармен қоныстастығын білдіретін мағынадағы туынды сын есімдер жасалса. Мыс.: *ауылдас, көршілес, жерлес, елдес, тумалас, ұялас, жапсарлас, іргелес* т.б.	223
	в) кейбір зат есімдерге жалғанып, олардың нақтылы лексикалық мағыналарына қарай, түрлі қарым-қатынас, мерзім-мөлшер жағынан сыбайлас келетін сындық ұғымдарды білдірсе. Мыс.: *жолдас, замандас, сабақтас, кәсіптес, қызметтес, дәмдес, өкшелес* т.б.	224
	8. ***-шаң, -шең*** жұрнақтары арқылы жасалатын туынды сын есімдер:	
	а) киім-кешек атауларына жалғанып, адамның бойындағы киімімен байланысты сыртқы бейне-көрінісін білдіретін сын есімдер жасалса. Мыс.: *көйлекшең, етікшең, пальтошаң, шалбаршаң* т.б.	225
	ә) кейбір зат есімдерге жалғанып, адамға не затқа тән белгілі бір ерекше қасиет білдіретін туынды сын есімдер жасалса. Мыс.: *ашушаң, сөзшең, бойшаң, тершең, кіршең* т.б.	226
2	9. Есімдерден туынды сын есім тудыратын өнімсіз жұрнақтар арқылы:	

4.3-кестенің жалғасы

1	2	3
	а) ***-дар, -дер, -тар, -тер*** қосымшалары кейбір зат есімдерге жалғанып, белгілі бір іс-әрекетіне душар болғандықты білдіретін сындық ұғым жасалса. Мыс.: *қарыздар, хабардар, борыштар* т.б.	227
	ә) иран тілдерінен (парсы, тәжік, ауған) ауысқан кірме ***"-и"*** қосымшасы кейбір есімдерге жалғанып, туынды сын есім жасалса. Мыс.: *әскери, мәдени, тарихи* т.б.	228
	б) ***-паз*** жұрнағы жалғанып, сын есімге бейім жаңа сөз жасалса. Мыс.: *өнерпаз, аспаз, білімпаз, ойынпаз* т.б.	229
	в) ***-мпаз (-м+паз), -ымпаз (-ым+паз), -імпаз (-ім+паз)*** жұрнағы (және компоненттер) кейбір есім сөздерге жалғанып, туынды сын есім жасалса. Мыс.: *сезімпаз, алымпаз, білімпаз, жағымпаз, жасампаз* т.б.	230
	г) ***-қой, (-гой, -гөй)*** жұрнағы кейбір зат есім есім сөздерге жалғанып, амал-әрекет пен мінез-құлықтардың атауы болатын туынды сын есім тудырса. Мыс.: *кәсіпқой, әзілқой, сәнқой, жәдігөй, әзілқой, сәнқой, жанжалқой* т.б.	231
	д) ***-қор*** жұрнағы зат есім есім сөздерге жалғанып, белгілі бір машықтанғанды білдіретін мағынадағы туынды сын есім жасалса. Мыс.: *жемқор, жалақор, айлақор, бейнетқор, қамқор, мансапқор, ызақор, намысқор* т.б.	232
	Етістіктерден өнімді жұрнақтар арқылы жасалатын туынды сын есім сөздер: 1. ***-қ, -к, -ық, -ік, -ақ, -ек:*** а) салт және сабақты етістікке жалғанып, түрлі сындық ұғымдағы атаулар жасайды. Мыс.: *ашық, тұнық, шірік, ілік, дөңгелек* т.б.; ә) еліктеу сөздерге жалғанып, қатыстық сын есім тудырса. Мыс.: *бұлтақ, жалтақ, еңкек, жалпақ* т.б.; б) кейбір зат есімдерге жалғанып, туынды сын есім жасаса. Мыс.: *жолақ, қасырақ, ортақ, ірек* т.б.; в) етістіктен ***-уық (-үік)*** формасы арқылы белгілі бір іс-әрекетке бейімділікті білдіретін туынды сын есім жасаса. Мыс.: *жылауық, сөйлеуік, сұрауық, сыбырлауық* т.б.	317
2	2. ***-ыңқы, -ңқы, -іңкі.*** Мыс.: *жатыңқы, салбыраңқы, көтеріңкі, кебіңкі, батырыңқы* т.б.	318

4.3-кестенің жалғасы

1	2	3
	3. ***-ынды, -інді, -нды, -нді.*** Мыс.: *асыранды, серпінді, жаттанды, шұбырынды, түйінді* т.б.	319
	4. ***-малы (-мелі, -балы, -белі, -палы, -пелі).*** Мыс.: *ауыспалы, көшпелі, таңдамалы, аумалы-төкпелі* т.б.	320
	5. ***-қыш, -кіш, -гыш, -гіш.*** Мыс.: *білгіш, оңғыш, тапқыш, сенгіш, айтқыш* т.б.	321
	6. ***-шақ (-шек).*** Мыс.: *мақтаншақ, еріншек, қызғаншақ, ашуланшақ, жасқаншақ, тартыншақ* т.б.	322
	7. ***-ымды (-імді, -мды, -мді).*** Мыс.: *жағымды, ұғымды, үйлесімді, жарасымды, сенімді* т.б.	323
	8. ***-улы, -улі.*** Мыс.: *жинаулы, ерттеулі, үюлі* т.б.	324
	9. ***-қақ, -кек, -ғақ, -гек.*** Мыс.: *асқақ, тоңғақ, жабысқақ, майысқақ, тайғақ, оңғақ, қатқақ, ұрысқақ* т.б.	325
	10. ***ма, -ме, -ба, -бе, -па, -пе.*** Мыс.: *жалдама (ақы), қызба (адам), көшпе (құм), сырма (бешпет), аспа (шам), сырма (бешпет)* т.б.	326
	Етістіктерден өнімсіз жұрнақтар арқылы жасалатын туынды сын есім сөздер: 1. ***-ыс, -іс, -с.*** Мыс.: *келіс, кетіс, ұқсас, таныс, тіркес, жалғас, талас* т.б.	327
	2. ***-қыр, -гыр, -кір, -гір.*** Мыс.: *білгір, ұшқыр, үшкір, алғыр, өткір, білгір* т.б.	328
	3. ***-мыс (-мыш).*** Мыс.: *жасамыс, айтылмыш, тарамыс, жазылмыш, жаралмыш* т.б.	329
	4. ***-ымтал (-імтал).*** Мыс.: *ұғымтал, өсімтал, сезімтал* т.б.	330
	5. ***-қы (-кі, -гы, -гі).*** Мыс.: *бұралқы, оралғы, жинақы, күлдіргі* т.б.	340
	6. ***-ыр (-ір, -ар, -ер, -р).*** Мыс.: *жұмыр, иір, обыр, қыңыр, тықыр, шымыр, құзар, былжыр, жалтыр, жылтыр, сылбыр, былбыр* т.б.	341
	7. ***-у.*** Мыс.: *жарау, қызу, тақау, жадау, жабырқау, тұзу, таяу, бітеу, баяу, көлбеу, қату* т.б.	342
	8. ***-ған (-еген).*** Мыс.: *тебеген, қабаған, сүзеген, қашаған, береген, алаған, безеген, көреген* т.б.	343
	9. ***-ын (-ін, -н).*** Мыс.: *ортан, бүтін, еркін, ұзын, жайын* т.б.	344
	10. ***-қалақ (-келек, -ғалақ, -гелек).*** Мыс.: *ұшқалақ, сасқалақ, қозғалақ* т.б.	345
2	11. ***-алақ (-елек).*** Мыс.: *еңкелек, қаңғалақ, бұлталақ, шыжалақ, жалталақ* т.б.	346
	12. ***-анақ.*** Мыс.: *сұғанақ, шұқанақ, шығанақ* т.б.	347
	13. ***-ғылықты (-гілікті, -қылықты, -кілікті).*** Мыс.: *жеткілікті, тұрғылықты, жергілікті* т.б.	348
	14. ***-мсақ (-мсек, -ымсақ, -імсек).*** Мыс.: *сұрамсақ, тілемсек, жарамсақ, берімсек, өлімсек* т.б.	349

Енді, осы *4.3-кестеде* қарастырылған сын есімнің лексика-морфологиялық құрылымына ақпараттық белгі-кодты сәйкестендіру бағдарламасы негізінде, кейбір негізгі және туынды сын есімдерге қатысты сөздерге де шартты белгі-код сәйкестендіру мысалдары *4.3.1* және *4.3.2-кестелерде* беріледі.

4.3.1-кесте

Сапалық (негізгі) сын есім сөздер	**Белгі-код орны**	
	1	**2**
қызыл, жасыл т.б.	CH1	201
жақсы, жаман т.б.	CH1	202
ұзын, қысқа, ауыр т.б.	CH1	203
тәтті, ащы т.б.	CH1	204

4.3.2-кесте

Қатыстық (туынды) сын есім сөздер	**Белгі-код орны**	
	1	2
сыртқы, төргі т.б.	CH2	210
кешкі, күндізгі т.б.	CH2	211
әсерлі, инабатты т.б.	CH2	212
үлкенді-кішілі, өзенді-сулы т.б.	CH2	213
қаз мойынды, теке сақалды т.б.	CH2	214
мұңсыз, білімсіз т.б.	CH2	215
ойшыл, өзімшіл т.б.	CH2	216
балғадай, әкедей, шөлмектей т.б.	CH2	217
қалалық, қоғамдық т.б.	CH2	218
айлық, жылдық т.б.	CH2	219
көйлектік, пальтолық т.б.	CH2	219
өздік, қаншалық т.б.	CH2	220
сырлас, ниеттес т.б.	CH2	221
туыстас, қарындас т.б.	CH2	222
ауылдас, көршілес т.б.	CH2	223
жолдас, қызметтес т.б.	CH2	224
көйлекшең, пальтошаң т.б.	CH2	225
ашық, шірік, дөңгелек т.б.	CH2	317
көтеріңкі, жатыңқы т.б.	CH2	318
жаттанды, серпінді т.б.	CH2	319
таңдамалы, көшпелі т.б.	CH2	320
білгіш, оңғыш, сенгіш т.б.	CH2	321
мақтаншақ, еріншек т.б.	CH2	322
жағымды, үйлесімді т.б.	CH2	323
жинаулы, ерттеулі т.б.	CH2	324
ұрысқақ, тоңғақ т.б.	CH2	325
қызба (адам), сырма т.б.	CH2	326
таныс, келіс т.б.	CH2	327
сұрамсақ, өлімсек т.б.	CH2	349

4.5. Сын есімнің синтаксистік тәсіл арқылы жасалу типтерін шартты белгі-кодпен сәйкестендіру бағдарламасы

Сын есім сөздер құрамына қарай, жалаң және күрделі болып екіге бөлінетіні айқын. Жалаң сын есімдер бір ғана компоненттен тұрады, бірақ онда қосымша морфеманың болуы я болмауы шарт емес.

Ал күрделі сын есімдерге екі, үш және одан да көп компоненттерден құралып барып, синтаксистік (аналитикалық) жолмен жасалған сын есім сөздер жатады. Күрделі сын есімдер негізгі сын есім сөздерден және олардың тіркесуі, қосарлануы, бірігуі арқылы жасалып, тілімізде бір бүтін күрделі тұлға ретінде қызмет ететіні белгілі.

4.4-кесте

Сын есімнің синтаксистік тәсіл бойынша жасалу құрылымына ақпараттық белгі-кодты сәйкестендіру бағдарламасы

Белгі-код орны	**Күрделі сын есімдер сипатының негізгі үлгілері**	**Белгі-код**
1	**Күрделі сын есім**	СН3
2	1. Сапалық (негізгі) сын есімдердің бірімен-бірі тіркеседі. Мысалы: *қара ала, сары ала, қызыл ала* т.б.; *қара көк, қара торы, қара күрең, қара шұбар* т.б.	400
	2. Бірыңғай я негізгі (сапалық), я туынды (қатыстық) сын есім сөздер не қайталанады, не қосарланады. Мысалы: *кішкене-кішкене, аппақ-аппақ, үлкен-үлкен* т.б.; *малды-малсыз, үлкенді-кішілі, қоралы-қопсылы* т.б.	401
	3. Негізгі сын есім сөзбен ***-лы, -лі*** жұрнақты туынды сын есім тіркеседі. Мысалы: *кең маңдайлы, қызыл шырайлы* т.б.; екі компоненті де не ***-лы, -лі*** не ***-сыз, -сіз*** жұрнақтары арқылы жасалған қос сөзден тұратын туынды (қатыстық) сын есімдер. Мысалы*: таулы-тасты, әкелі-балалы, ессіз-түссіз, жөнді-жөнсіз* т.б.	402
	4. Зат есім немесе сан есім сөздермен ***-лық (-лік, -дық, -тік, -тық, -тік)*** жұрнақтар арқылы жасалатын туынды сын есімдер тіркесіп, күрделі сын есім жасалады. Мысалы: *халық аралық, бес жылдық, екі кісілік, екі-үш күндік* т.б.; *көңіл көтererлік, қабырға қайысарлық* т.б.	403
	5. Екі компоненттің екеуіне де бірдей не ***-лық, -лік, -дық, -дік, -тық, -тік, -и*** не бір компонентті ***-н*** жұрнақтары арқылы жасалған күрделі сын есімдер. Мысалы: *әскери-саяси, ғылыми-әдеби, қоғамдық-тарихи, әлеуметтік-экономикалық, статистикалық-лингвистикалық* т.б.	404
	6. Тәуелділіктің 3-ші жағының қосымшасы қосылған зат есімге сын есім тіркесіп, күрделі сын есім есебіне қолданылады. Мысалы: *көзі ашық, тілі майда, басы бос, қолы қысқа, жүзі жылы* т.б.	405

Мәтін ішінде кездесетін сын есімнің синтаксистік тәсілі бойынша жасалатын сөздерді шартты белгілермен белгілеу немесе тиісті белгі-кодтпен сәйкестендіру жоғарыда көрсетілген *4.4-кесте* арқылы іске асуы қажет деп ұйғардық.

Аталған кестедегі мәліметтерді анықтай түсу үшін төменде сәйкестіктердің мысалы ретінде *4.4.1-кестені* ұсынып отырмыз. Мәтіндегі не жиілік сөздіктегі сын есімнің синтаксистік тәсіл арқылы жасалған түрлерін осылайша шартты түрде белгілеу негізінде оларға зерттеу кезінде ықтималды-статистикалық әдісті қолданудың мүмкіндігі туады.

4.4.1-кесте

Күрделі сын есім сөздер	**Белгі-код орны**	
	1	**2**
қара ала, сары ала; қара көк, қара торы, қара күрең т.б.	СН3	400
кішкене-кішкене, аппақ-аппақ, үлкен-үлкен т.б.; *малды-малсыз, үлкенді-кішілі, қоралы-қопсылы* т.б.	СН3	401
кең маңдайлы, қызыл шырайлы т.б.; *таулы-тасты, әкелі-балалы, ессіз-түссіз* т.б.	СН3	402
халық аралық, бес жылдық, екі кісілік, екі-үш күндік т.б.; *көңіл көтererлік, қабырға қайысарлық* т.б.	СН3	403
әскери-саяси, ғылыми-әдеби, қоғамдық-тарихи, әлеуметтік-экономикалық, статистикалық-лингвистикалық т.б.	СН3	404
көзі ашық, тілі майда, басы бос, қолы қысқа, жүзі жылы т.б.	СН3	405

Сонымен, осы тарауда қарастырылған қазақ тілінің негізгі сөз таптары ішінен зат есім, етістік және сын есімге қатысты сөз бен сөзтұлғаларды шартты түрде белгілеу, яғни олардың түр-түріне белгі-код сәйкестендіру бағдарламасы қазақ тілінің морфологиялық құрылымына статистикалық әдісті молынан қолдануға мүмкіндік тудырады. Мұндай бағдарламаны қазақ тілінің басқа да сөз таптары бойынша, оларға тән ерекшеліктерін ескере отырып, баяндалған жолмен іске асыруға болатындығына сеніміміз мол. Біз ұсынып отырған бағдарлама үлгісі әзірше теориялық ізденістердің бірі болғандықтан тәжірибе жүзінде толығынан қолданыс таба қойған жоқ. Бірақ мұндай статистикалық зерттеулер қазақ тілі үшін көп күттірмейтін жуық болашақтың ісі болмақ.

ҚОРЫТЫНДЫ

Компьютер көмегімен алынған қазақ тілінің көркем әдебиет және басқа да стильдер мәтіндерінен түзілген жиілік сөздіктерден (әліпбилік, жиілік, кері-әліпбилік, сөзнұсқағыш) лингвостатистикалық зерттеу арқылы алынған деректер қазақ мәтіндерін автоматты түрде өңдеуде, ұзақ мерзімге сақтауда және қажетті мәліметтерді іздеп, табуда аса қажет.

М.О.Әуезовтің қаламынан шыққан "Абай жолы" романының мәтіні тұңғыш рет статистикалық зерттеу нысанына айналып, көркем әдебиеттің бас жиыны (генералды жиын) немесе ықтималдық нысанының бас нұсқасы ретінде қазақ мәтінінің "электронды-ықтималдық сөздігінің" статистикалық үлгісіне айналды.

Өзге тілдерді зерттеуші ғалымдардың тұжырымдауы бойынша, бірқатар статистикалық сипаттамалар көптеген тілдерге ортақ болса, енді бірқатары тілдің өзіндік ерекшелігіне қарай, жанрлық айырымына, автордың тіліне ғана тән ерекшеліктеріне байланысты әртүрлі болады. Қазақ тілі мәтіндерінің де өзіне тән ерекшеліктері баршылық. Мәселен, әртүрлі стильдер мәтіндерінен түзілген жиілік сөздіктердегі сөзтұлғалар жиілігінің мәтінді не сөздікті қамту сипаты бірыңғай емес.

Түркі тілдеріндегі мәтіндерді статистикалық әдіспен зерттеудің теориялық та маңызы зор. Агглютинативтік құрылымды түркі тілдеріндегі зерттеулер нәтижелері флективті-аналитикалық құрылымдағы үндіеуропа тілдерінің бай статистикалық деректерімен салыстыра қарастыруға мүмкіндік береді. Мысалы мынандай жайтқа көңіл бөлейік. Орыс, ағылшын, неміс, француз, румын, испан және т.б. тілдерде кез келген мәтіннің 50 пайызын қамту үшін ең жоғары жиілікті 100–150 сөзтұлға жеткілікті болса, ал түркі тілдерінде – жоғары жиілікті 700–800 сөзтұлға қажет екен.

Түркі тілдері бойынша ақпараттық сипаттама мөлшері де үндіеуропа тілдеріне қарағанда басқаша болуы мүмкін. Сондықтан лингвостатистикалық зерттеулер әрі типологиялық эксперимент аясына жатады деп түсінген жөн.

Еңбекте қазақ мәтіндері негізінде орындалған ықтималды-статистикалық және ақпараттық үлгі жасалып, олардың мүмкіндіктері белгіленді. Осындай зерттеу жұмысының нәтижелері бойынша, жиілік сөздіктің бастапқы жағынан соңына қарай қарастырғанда, сөзтұлғалардың мәтін бойын статистикалық және ақпараттық тұрғыда қамту сипатының өзіне тән ерекшеліктері тиісті теориялық критерийлер арқылы анықталды. Қазақ тілі мәтіндері бойынша түзілген жиілік сөздіктердегі сөзтұлғалардың көп жағдайда мәтінге тәуелсіз-ақ мағыналық дәрежеде болатындығы айқындалып отыр.

Қорыта айтқанда, кейбір тілдік құбылыстар өте әлсіз көрініс тауып, жасырын сипатта тұрып, көп жағдайда зерттеушінің тікелей бақылауынан тыс қалады. Міне, осындай тілдік құбылыстарды анықтайтын математикалық құралды статистикалық зерттеулер арқылы табуға болады. Ондай құбылыстарға кейбір сөз таптарына қатысты *сөз, сөзтұлға, сөз тіркестердің* және т.б. бірліктердің тілдегі қолдану ерекшеліктері жатады.

Зерттеуімізде аталған математикалық құрал ретінде алынғандар:

а) айнымалы шаманың (тілдік бірліктер жиілігі) мәтін ішінде үлестірілуінің статистикалық заңдылықтары;

ә) тілдік бірліктердің статистикалық және ықтималдық сипаттарының арақатынастарын бағалайтын теориялық критерийлер.

Сонымен аталмыш оқу құралы түркі тілдеріндегі, соның ішінде қазақ тіліндегі мәтіндердің құрылымын объективті түрде квантитативтік әдіс-тәсілдермен зерттеуде көмегі болатындығына сеніміміз мол.

ҚОСЫМША

Қосымша 1

Оқу құралында пайдаланылған қазақша-орысша терминдер сөздігі

абсолютті ауытқу – абсолютное отклонение
абсолютті жиілік – абсолютная частота
абсолютті қате – абсолютная ошибка
абсцисс өсі – ось абсцисс
автоматты құрылғы – автоматическое устроиство
автоматты лингвистика – автоматическая лингвистика
автоматтандырылған түр – автоматизированный вид
айырым модулі – модуль разности
айнымалы шама – варирующая величина
ақиқат оқиға – достоверное событие
ақиқаттық аралық мән – действительное интервальное значение
ақпарат – информация
ақпаратты автоматты өңдеу – автоматическая обработка информации
ақпаратты автоматты іздеу (тауып алу) – автоматический поиск информации
ақпараттық өлшем – информационное измерение
ақпаратты сақтау – хранение информации
ақпаратты жолдау (тарату) – передача информации
ақпаратты ұзақ мерзімге сақтау – длительное хранение информации
ақпараттық артықтық (энтропия) – информационная избыточность
ақпараттық жүйе – информационная система
ақпараттық салмақ – информационный вес
ақпараттық салмақтың өсу қарқыны – темп (скорость) возрастания информационного веса
ақпарат теориясы – теория информации
ақпараттық үлгі (модель) – информационная модель
алгоритм – алгоритм
алшақтық дәрежесі – степень расхождения
алшақтық сипаты – характер расхождения
алшақтықты бағалау (сынау) – оценка расхождения
альтернативті (қарсы қойылатын) болжам – альтернативная гипотеза
аналитикалық (сараптамалық) түрпат – аналитическая форма
аралық код – промежуточный код
арифметикалық орта – средняя арифметическая
ауытқу шамасы – величина отклонения
ауытқудың орта шамасы – среднее значение от отклонений
ауытқулардың абсолют шамаларының ортасы – среднее от абсолютных отклонений
ашық жүйе – открытая система
әдіс – метод
әліпби-жиілік сөздік – алфавитно-частотный словарь
әліпби ретімен (бойынша) – по алфавиту
әліпбилік құрам – алфавитный состав
әмбебаптық сипат – универсальный характер
бағдарлама жұмысы – работа программы

базалық (негізгі) тіл – базовый язык
байқау саны – число наблюдений
байланыс күші – сила связи
байланыс торабы – узел связи
байланыс тығыздығы – теснота связи
бақылаулар тізбегі – последовательность испытаний
бақылау қатесі – ошибка наблюдения
бақылау (тәжірибе) саны – число испытаний
бас (генералды) мәтін жиынтығы – генеральная совокупность текстов
белгі (таңба) – знак
белгі-код – знак-код
белгілер жүйесі – система знаков
белгілеу (кодтау) әдісі – метод (прием) кодирования
билогарифмдік координат – билогарифмический координат
болжам – гипотеза
болжамды бағалау (тексеру) – оценка (проверка) гипотезы
болмыс бөлігі – часть действительности
бос аралық – пробел
буындық құрылым – слоговая структура
білім жүйесі – система знаний
білімді модельдеу (үлгілеу) – моделирование знаний
біртекті – однородный
біртектілік – однородность
бірлік – единица
бірліктер сипаттамасы – характеристика единиц
бірліктер арақатынасы – соотношение единиц
бірізділік (стандарт) – стандарт
генералды (бас) жиын (жиынтық) – генеральная совокупность
геометриялық орта – средняя геометрическая
грамматикалық ақпарат – грамматическая информация
грамматикалық белгі – грамматический признак
грамматикалық категория – грамматическая категория
грамматикалық көрсеткіш – грамматический показатель
ғылыми ақпарат – научная информация
дедуктивті – дедуктивный
дискретті вариациялық үлестірілу қатары – дискретный вариационный ряд распределения
дискретті бірліктер тізбегі – последовательность дискретных единиц
дисперсия (орта квадраттық ауытқу) – дисперсия
екінші (үшінші) реттегі орталық момент – центральный момент второго (третьего) порядка
елеусіз (сипаттағы) ауытқу – несущественное отклонение
елеулі (сипаттағы) ауытқу – существенное отклонение
ең кіші квадраттар әдісі – метод наименьших квадратов
еркіндік дәреже саны – число степеней свободы
есептеу лингвистикасы – вычислительная лингвистика
жабық жүйе – закрытая система
жасанды интеллект – искусственный интеллект
жасанды тіл – искусственный язык
жиынтық абсолюттік жиілік – накопленная абсолютная частота
жиынтық ақпараттық салмақ – накопленный информационный вес
жиынтық қатынастық жиілік – накопленная относительная частота
жиілік – частота
жиілік сөздік – частотный словарь

жиілікті сөзтізбе – частотный словник
жиіліктердің үлестірілу заңдылығы – закон распределения частот
жоғары математика – высшая математика
жүйе – система
зерттеу нысаны – объект исследования
инженерлік лингвистика – инженерная лингвистика
индуктивті – индуктивный
интервал – интервал
интервалдық бөлік – интервальная часть
картотекалық қор – картотечный фонд
квадраттанған үлес ауытқуы – квадратичное отклонение доли
квантитативті құрылым – квантитативная структура
квантитативті лингвистика – квантитативная лингвистика
кері тәуелділік – обратная зависимость
кездейсоқ алшақтық – случайное расхождение
кездейсоқ жағдай – случайное обстоятельство
кездейсоқ лингвистикалық шама – случайная лингвистическая величина
кездейсоқ құбылыс – случайное явление
кездейсоқ оқиға (жағдай) – случайное событие
кездейсоқ таңдама – случайная выборка
кездесу жиілігі – частота встречаемости
кездесу ықтималдығы – вероятность встречаемости
келісім критерийі – критерий согласия
кему тәртібі – порядок убывания
кері әліпби-жиілік сөздік – обратно-алфавитный частотный словарь
кері пропорционал шама – обратно-пропорциональная величина.
кері сөздік – обратный словарь
кесте деректері – табличные данные
күткен оқиға – ожидаемое событие
кодтау принципі – принцип кодирования
комбинаторлық қасиет – комбинаторное свойство
компьютер – компьютер
компьютерлік аударма – компьютерный перевод
компьютерлік бағдарлама – компьютерная программа
компьютерлік қор – компьютерный фонд
компьютерлік лингвистика – компьютерная лингвистика
координат өстер жазықтығы – плоскость координатных осей
корреляция (өзгермелі шамалар арасындағы тәуелділік) – корреляция
корреляция коэффициенті – коэффициент корреляции
корреляциялық талдау – корреляционный анализ
кілт мәні (шамасы) – ключевое значение
қамту қарқыны – скорость (темп) покрываемости
қамту пайызы – процент покрываемости
қалыптастыру – формирование
қатынастық жиілік – относительная частота
қатынастық қате – относительная ошибка
қатынастық сипат – относительный характер
қатынастық-функционалдық салмақ – относительно-функциональный вес
қиылысу нүктесі – точка пересечения
қолайлы жағдай – благоприятствующий случай
қолайсыз жағдай – неблагоприятствующий случай
қолданбалы лингвистика – прикладная лингвистика
қолдану жиілігі – частота употребления
құбылу – варьирование (изменчивость)

құбылу (вариация) коэффициенті – коэффициент вариации
құжат – документ
құжат өрісі – поле документа
құрылымды-грамматикалық – структурно-грамматическая
лексика-морфологиялық құрылым – лексико-морфологическая структура
лексикалық белгі – лексический признак
лексикалық қор – лексический фонд
лингвистика (тіл білімі) – лингвистика
лингвистикалық бірлік – лингвистическая единица
лингвистикалық болжам – лингвистическая гипотеза
лингвистикалық мағына – лингвистическое значение
логика-математикалық – логико-математическая
маңыздылық ақпарат (семантикалық) – значимая информация (семантическая)
маңыздылық деңгей – уровень значимости
мағыналық жан-жақтылық – разнообразность значений
мазмұнды формальдау – формализация содержания
максимумдық шама – максимальное значение
математикалық баға – математическая оценка
математикалық лингвистика – математическая лингвистика
математикалық өрнек – математическая формула
математикалық статистика – математическая статистика
маңыздылық деңгей – уровень важности
машиналық аударма – машинный перевод
мәнмәтін – контекст
мәнмәтіндік қоршау – контекстное окружение
мәтін – текст
мәтін көлемі– объем текста
мәтін лингвистикасы – лингвистика текста
мәтін туындауы – порождение текста
мәтін ұзындығы – длина текста
мәтінді қамту – покрываемость текста
мәтінді нормалау – нормирование текста
мәтіндік бірлік – текстовая единица
мәтіннің семантикасы – семантика текста
мәтіндерді автоматты өңдеу – автоматическая обработка текстов
микротаңдама – микровыборка
микрожүйе (ең кіші жүйе) – микросистема
минимум сөздік – словарь минимум
модельденуші мәтін – моделирующий текст
морфемдік құрам – морфемный состав
морфологиялық белгі – морфологический признак
морфологиялық деңгей – морфологический уровень
морфологиялық (синтаксистік, семантикалық) тармақтану – морфологическая (синтаксическое, семантическая) разветвление
морфологиялық түлға – морфологическая форма
мүмкін емес оқиға – невозможное событие
нағыз орта жиілік – действительная средняя частота
нағыз үлес – действительная доля
нақты ғылымдар – точные науки
негізгі нысан – основной объект
нормалау сипаты – характер нормирования
нормалық табиғат (тілдің) – природа нормы (языка)

нормальды үлестірілу – нормальное распределение
нөлдік болжам – нулевая гипотеза
нүктелер жүйесі – система точек
объективті – объективный
оқига – событие
оқиганың заңдылыгы – закономерность события
оқиганың ықтималдыгы – вероятность события
оқиганың шыгуы (пайда болуы) – появление события
ординат өсі – ось ординат
орта ақпарат мәні – среднее информационное значение
орта ақпарат саны – среднее информационное число
орта жиілік – средняя частота
орта жиіліктен ауытқу – отклонение от средней частоты
орта квадраттық ауытқу – среднее квадратичное отклонение
орта таңдама жиілік – средняя выборочная частота
орта шамадан ауытқу – отклонение от средней величины
орта энтропия – средняя энтропия
өзгермелі сипат – изменчивый характер
өлшенген ақпарат – взвешенная информация
өрнек – выражение (формула)
өспелі функция – возрастающая функция
пайыздық деңгей – процентный уровень
параметрлерді нормалау – нормирование параметров
ранг (реттік нөмір) – ранг (порядковый номер)
регрессия түзуі – прямая регрессии
рет саны – порядковый номер
салыстырма деректер – сравнительные данные
санау құралы – инструмент подсчета
сандық деректер – числовые данные
сандық көрсеткіш – количественный показатель
сандық қатынас – количественное соотношение
сандық сипаттама – количественная характеристика
салыстырмалы-типологиялық – сравнительно-типологическая
селективті ақпарат – селективная информация
семантикалық белгі – семантический признак
семантикалық көрініс – семантическое представление
сенімділік – надежность
сенімділікті бағалау (тексеру) – оценка надежности
сериялық бөліктеу – деление на серий
символ – символ
символдар жүйелілігі – системность символов
синтаксистік белгі – синтаксический признак
синтаксистік бірлік – синтаксическая единица
синтаксистік деңгей – синтаксический уровень
синтаксистік қызмет – синтаксическая функция
синхрондық лингвистика – синхронная лингвистика
сипаттау грамматикасы – описательная грамматика
сөз байлық – словесное богатство
сөз статистикасы – статистика речи
сөздік бірлігі – единица словаря
сөздік қор – словарный фонд
сөздік құрылымы – структура словаря
сөздік өрісі – словарное поле

сөздік үзіндісі – отрывок словаря
сөздікті қағазға шығару – распечатка словаря
сөздікті қамту – покрываемость словаря
сөзқолданыс – словоупотребление
сөзнұсқағыш – словоуказатель
сөзнұсқағыш әліпби-жиілік сөздік – алфавитно-частотный словарь словоуказатель
сөзтіркес – словосочетание
сөзтұлға – словоформа
сөзформа (сөзтұлға) – словоформа
сөйлеу ақиқаттығы – речевая действительность
сөйлеу арнасы (байланыс арнасы) – канал речи (канал связи)
статистика-ақпараттық – статистико-информационная
статистикалық әдіс – статистический метод
статистикалық бағалау – статистическая оценка
статистикалық байланыс теориясы – статистическая теория связи
статистикалық бояу – статистическая окраска
статистикалық деректер – статистические данные
статистикалық заңдылық – статистическая закономерность
статистикалық құрал – статистический инструмент
статистикалық құрылым – статистическая структура
статистикалық лингвистика – статистическая лингвистика
статистикалық мәліметтер – статистические факты
статистикалық санақ – статистический подсчет
статистикалық сипаттама – статистическая характеристика
статистикалық үлестірілу – статистическое распределение
стильдік ерекшелік – стилевое отличие
стильдік-семантикалық – семантико-стилистическая
субъективті – субъективный
сызба – график
сызба-топтама – блок-схема
сызықтық пішін – прямолинейная форма
сызықтық функция – линейная функция
сынақ таңдамалары – оценочные выборки
сынама үлгі – испытательная модель
сынық сызықтық түрпат – ломанная прямая
сынық қисық – ломанная кривая
таңба – знак
таңдама бөлік – выборка
таңдама жиілік – выборочная частота
таңдама мәтіндер – выборочные тексты
таңдама жиіліктің орта жиіліктен ауытқуы – отклонение выборочных частот от среднего (значения)
тәуелді модельдер – зависимые модели
тәжірибе – эксперимент (испытание)
тәжірибе үстінде – во время испытания
тәжірибелік жиіліктер үлестірілуі – распределение эмпирических частот
теориялық-жиынтық – теоретико-множественная
теориялық-ақпараттық зерттеу – теоретико-информационное исследование
теориялық жиіліктер үлестірілуі – распределение теоретических частот
теориялық коэффициент – теоретический коэффициент
теориялық қатынастық жиынтық жиілік – теоретическая относительная накопленная частота
теориялық сызық – теоретическая линия

теориялық үлестірілу заңдылықтары – теоретический закон распределения
теориялық үлестірілу – теоретическое распределение
термин (атау) – термин
терминтану – терминоведение
техникалық және іскерлік коммуникация – техническая и деловая коммуникация
топтама – блок
түзу сызықтық байланыс тығыздығы – теснота прямолинейной связи
түлға – форма
тұрақты коэффициент – постоянный коэффициент
тұрақты сан (константт) – постоянное число (константа)
тұрақты шама – постоянная величина
тұрақтылық сипат – постоянный характер
тік сызықтық пішін – прямолинейный вид
тікелей байқау – непосредственное наблюдение
тікелей құрастырушылар моделі – модель непосредственно составляющих
тікелей сызықтық регрессия – прямолинейная регрессия
тікелей тәуелділік – прямая зависимость
тіл арқылы қатынас жасау (тілдік қатынас) – речевое общение
тілдесу – диалог
тілді ақпараттық бағалау – информационная оценка языка
тілдік ақпарат – языковая информация
тілдік бірлік – языковая единица
тілдік дерек – языковое данное
тілдік элемент – языковой элемент
тілдің ықтималдық үлгісі (моделі) – вероятностная модель языка
тірек сөз – опорное слово
үздіксіз вариациялық қатар – непрерывный вариационный ряд
үздіксіз шама – непрерывная величина
үйлесімдік дәреже – степень согласия
үйлесімдік дәрежені бағалау – оценка степени согласия
үйлесімдік критерийі – критерий согласия
үлгі (модель) – модель
үлес – доля
үлес ауытқуы – отклонение доли
үлестің тербелуі – колебание доли
үлестірілу (таралу) – распределение
үлестірілу түрпаты – форма распределения
үлестірілудің қисық сызықтық пішіні – положение кривой распределения
ұя – ячейка
ұялар топтамасы – блок ячеек
ұялар тізбегі – последовательность ячеек
функционалдық мән (мағына) – функциональное значение
функционалдық салмақ – функциональный вес
фонемалық спектр – фонемный спектр
фонемалық топ – фонемная группа
фонетикалық құрылым – фонетическая структура
формальды әдіс – формальный метод
формальды модель (формальды үлгі) – формальная модель
формула (өрнек) – формула
хабарламаны қабылдау – прием сообщения
хабарламаны тарату – передача (распространение) сообщения
шағын тіл – подъязык
шартты код – условный код
шартты кодтарға сәйкестендіру – ставить в соответствие условный код

шартты кодтау ұстанымы – придерживаться кодификации
шашырамдық дәреже – степень разбросанности
шекті дискретті бірлік – ограниченная дискретная единица
шектеулерді анықтау және бағалау – определение и оценка ограничений
шындық мән (мағына) – истинное значение
ықтималды-статистикалық әмбебап сызба – универсальная вероятностно-статистическая схема
ықтималды-статистикалық заңдылық – вероятностно-статистическая закономерность
ықтималды-статистикалық үлгі (модель) – вероятностно-статистическая модель
ықтималды-үлестірімді заңдылық – вероятностно-распределительная закономерность
ықтимал оқиға – вероятное событие
ықтималдық интеграл – интеграл вероятностей
ықтималдық қатынастар – вероятностные отношения
ықтималдық шама – вероятностная величина
ықтималдықтар теориясы – теория вероятностей
ЭЕМ – ЭВМ
электронды-есептеу машинасы – электронно-вычислительная машина
электронды-ықтималдық сөздік – электронно-вероятностный словарь
элемент – элемент
эмпирикалық сызық – эмпирическая линия
эмпирикалық (тәжірибелік) үлестірілу – эмпирическое распределение
энтропия – энтропия

Қосымша 2

Қолданбалы лингвистика мен математикалық лингвистика пәндерінде қолданылатын негізгі терминдердің орысша-қазақша сөздігі және анықтамалары

1. Алгоритм – *алгоритм.* Нақты мақсатқа жету үшін орындалатын амалдар реттілігі (тізбегі).

2. База данных – *деректер қоры.*

Ақпаратты автоматты түрде тарату, түсініктеме беру мен өңдеудің формальды көрінісінің қолайлы түрі.

3. База знаний – *білімдер қоры.*

Тілден сырт жатқан ақиқат шындық бөлшектері жайлы ілімнің формальды түсініктерінің элементтері.

4. Байт – *байт.*

Есептеу техникасындағы бірлік. Әдетте, біртұтас бірлік түрінде қарастырылатын екілік разрядтың тізбегі (көбінде тізбек саны 8-ге тең). Есептеу техникасы жадының өлшемі: 1 килобайт =1024 байт, 1мегабайт =1024 килобайт. Мәтінді өңдеу мен оны жадында сақтау жағдайында бір байт – бір символға сәйкес келеді.

5. Грамматика зависимостей – *тәуелділік грамматикасы.*

Құрамдастар араларында иерархиялық тәуелділігі бар сөйлем құрылымының формальды берілісі. Иерархия – төменгі дәрежедегі элементтердің жоғарғы дәрежедегілерге бағыну тәртібі.

6. Грамматика непосредственно составляющих – *тікелей құрастырушылар грамматикасы.*

Барынша тәуелсіз, сызықтық пішіндегі бір-бірімен қиылыспайтын элементтердің иерархиялық түрде біріне-бірі енетін сөйлем құрылымының формальды көрінісі.

7. Грамматика представлений – *түсініктер (көріністер) грамматикасы.* Әрбір сөздің екі жақтық (сол және оң жақ) қоршалу мүмкіндіктері, яғни тіркесу мүмкіндігіне негізделген ережелер жүйесі.

8. Граф – *граф.*

Көп төбелі (нүктелі) және көп қабырғалы (байланысты) жиынның қос төбелерін қосатын (қосақтап) математикалық нысана.

9. Дерево предложения, дерево зависимостей – *сөйлем тармағы, тәуелділік тармағы.*

Тәуелді құрылымдардың құрамдас түйіндеріне сәйкестікке бейімделген граф түріндегі сөйлем құрылымының түсінігі.

10. Интерфейс – *интерфейс.*

Ақпараттар алмасуының амалдары мен тәсілдері. Адам мен машина аралығындағы интерфейс – адам мен ЭЕМ арасында диалогты ұйымдастырудың амалдары мен тәсілдері. Есептеу машиналары, бағдарламалар мен олардың жеке блоктары арасында да интерфейстер болады. Жасанды интеллектіге қатысты тілдік мәселенің бірі – табиғи-тілдік интерфейс құрастыру, яғни адамның ЭЕМ-мен (компьютермен) табиғи тілде сөйлесу мүмкіндігін іске асыру.

11. Квазиреферат – *квазиреферат.* Компьютер арқылы құрастырылған реферат (қысқаша мазмұн).

12. Конфигурационный анализ – *пішін үйлесімін талдау.*

Компьютерге енетін мәтінді екінші тілге аудару қажеттігіне қатысты тілдің алдын ала белгіленген синтаксистік "пішін үйлесімдер" жиыны негізінде талдау жүргізу. Егер компьютерге енетін мәтіннен жасалған "пішін үйлесімдігі" алдын ала белгілі "пішін үйлесімдігімен" сәйкес келсе, мәтіндік "пішін үйлесімдік" танылды деп саналады да, ол "шағымдалған пішінге" ие болып, әрі қарай да осындай ұстаныммен талданады. Жинақтау (синтездеу) кезінде талдау арқылы алынған осындай "пішін үйлесімділік" аударылуға тиісті тілдің "пішін үйлесімділігімен" салыстырылады. "Пішін үйлесімділік" сәйкестіктер алдын ала белгіленеді.

13. Лингвистический процессор – *лингвистикалық процессор.* Автоматтандырылған жүйелер үшін ЭЕМ-ге енетін табиғи тілдегі мәтіндерді өңдеу амал-әрекеттерінің жиынтығы.

14. Накопитель – *жинақтаушы.* Сыртқы құрылғы арқылы ақпаратты оқитын және жазатын ЭЕМ-нің (компьютердің) қондырғысы. Мысалы, магнит дискідегі жинақтаушы (дисковод), магнит таспадағы жинақтаушы.

15. Оконный интерфейс – *терезелі интерфейс.*

Соңғы кездегі интерфейс құру тәсілдерінің жетістігі болып саналады. Бұл тәсіл бойынша компьютер экранында қажетті ақпарат ерекше көрініс табады. Пішіні төрт бұрышқа тең ая, немесе "терезе" арқылы зерттеуші қажеттілікке сай компьютер жадына "енетін" және "шығатын" ақпараттарды орналастырады. Тілдесу (диалог) кезінде ондай "терезелердің" саны қажеттікке сай өсіп отырады.

16. Предсказуемостный анализ – *болжамдық талдау.*

Мәтін ішіндегі сөзқолданыстарды солдан оңға қарай талдай отыра зерттеу әрекеті (әдіс-тәсіл). Мәтіннен қарастырылатын әрбір келесі сөзқолданыстың алдын ала болжауға алынған сөзбен сәйкестігі тексеріледі.

17. Семантическое представление данных – *деректердің семантикалық берілісі.*

Семантикалық модельді сипаттауда қолданыс табатын ақпарат түсінігінің мазмұнын формальды түрде жазу тәсілі.

18. Словарь-конкорданс – *конкорданс-сөздік.*

Әрбір сөзтұлғаның көптеген мәтіндерде қолдану мүмкіндігіне сілтеме жасау үшін арнайы түзілген сөздік.

19. Тезаурус – *тезаурус.*

Тілдік бірліктер арасындағы семантикалық қатынасты көрсететін идеографикалық сөздік (идеография – әрбір таңбамен кез келген сөзді белгілеу). Тезаурустың құрылымдық негізі – нақты пәндік саланың иерархиялық ұғымдар жүйесі.

20. Файл – *файл (дерекжиын).* Біртұтас ретінде қарастырылатын компьютер жадындағы өзара байланыстағы жазулар жиыны.

21. Фрейм – *фрейм.* Мүмкін болатын құрамдар мен олардың арасындағы қатынас түрлерін атап көрсету үшін қажетті мәтін құрылымын не ситуацияны сипаттайтын амал.

22. Язык представления данных – *деректерді (берілістерді) сипаттау тілі.*

Ақпаратты оның құрылымы арқылы жазудың формальды әдісі. Деректерді сипаттаудың қайсыбір формальдық (математикалық) моделіне сүйенеді.

Қосымша 3

М.Әуезовтің "Абай жолы" роман-эпопеясының жиілік сөздігінен үзінді (ең жиі қолданылған 500 сөз)

Рет саны (i)	СӨЗ	Сөз табы	Абсолютті жиілік	Жиынтық абсолютті жиілік	Қатынастық жиілік	Жиынтық қатынастық жиілік
1	2	3	4	5	6	7
1	де	ет	9828	9828	0,02110	0,02110
2	бол	ет	9341	19169	0,02006	0,04116
3	е	ет	7844	27013	0,01684	0,05800
4	Абай	зт	6747	33760	0,01449	0,07249
5	өз	ес	5747	39507	0,01234	0,08483
6	да	шл	5653	45160	0,01214	0,09697
7	кел	ет	5175	50335	0,01111	0,10808
8	бұл	ес	4696	55031	0,01008	0,11816
9	ал	ет	4546	59577	0,00976	0,12792
10	де	шл	4506	64083	0,00968	0,13760
11	осы	ес	4379	68462	0,00940	0,14700
12	айт	ет	4301	72763	0,00924	0,15624
13	сол	ес	4175	76938	0,00897	0,16521
14	ол	ес	4117	81055	0,00884	0,17405
15	бір	са	3896	84951	0,00837	0,18242
16	отыр	ет	3568	88519	0,00766	0,19008
17	сөз	зт	2913	91432	0,00626	0,19634
18	қал	ет	2685	94117	0,00577	0,20211
19	бер	ет	2588	96705	0,00556	0,20767
20	жоқ	мд	2554	99259	0,00548	0,21315
21	үй	зт	2535	101794	0,00544	0,21859
22	тұр	ет	2479	104273	0,00532	0,22391
23	кет	ет	2451	106724	0,00526	0,22917
24	бар	мд	2304	109028	0,00495	0,23412
25	мен	шл	2261	111289	0,00486	0,23898
26	екі	са	2254	113543	0,00484	0,24382
27	көп	мд	2226	115769	0,00478	0,24860
28	ет	ет	1954	117123	0,00420	0,25280
29	шық	ет	1946	119669	0,00418	0,25698
30	күн	зт	1890	121559	0,00406	0,26104
31	енді	үс	1858	123417	0,00399	0,26503
32	жүр	ет	1838	125255	0,00395	0,26898
33	көр	ет	1830	127085	0,00393	0,27291
34	ел	зт	1793	128878	0,00385	0,27676
35	ауыл	зт	1726	130604	0,00371	0,28047
36	бір	ес	1708	132312	0,00367	0,28414
37	қара	ет	1586	133898	0,00341	0,28755
38	бала	зт	1545	135443	0,00332	0,29087
39	бар	ет	1512	136955	0,00325	0,29412
40	бас	зт	1491	138446	0,00320	0,29732

1	2	3	4	5	6	7
41	ғана	шл	1484	139930	0,00319	0,30051
42	көз	зт	1472	141402	0,00316	0,30367
43	ат	зт	1464	142866	0,00314	0,30681
44	үлкен	сн	1452	144318	0,00312	0,30993
45	не	ес	1425	145743	0,00306	0,31299
46	іші	зт	1425	147168	0,00306	0,31605
47	біл	ет	1387	148555	0,00298	0,31903
48	бірақ	шл	1364	146619	0,00293	0,32196
49	сал	ет	1346	151265	0,00290	0,32486
50	Құнанбай	зт	1293	152558	0,00278	0,32764
51	жай	зт	1263	153821	0,00271	0,33035
52	тағы	шл	1232	155053	0,00265	0,33300
53	жер	зт	1220	156273	0,00262	0,33562
54	жақ	зт	1195	157468	0,00257	0,33819
55	ғой	шл	1177	158645	0,00253	0,34072
56	сен	ес	1123	159768	0,00241	0,34313
57	қатты	сн	1088	160856	0,00234	0,34547
58	қой	ет	1085	161941	0,00233	0,34780
59	баста	ет	1047	162988	0,00225	0,35005
60	қол	зт	1047	164035	0,00225	0,35230
61	адам	зт	997	165032	0,00214	0,35444
62	жат	ет	993	166025	0,00213	0,35657
63	жол	зт	984	167009	0,00211	0,35868
64	кісі	зт	976	167985	0,00209	0,36077
65	және	шл	963	168948	0,00207	0,36284
66	жет	ет	912	169860	0,00196	0,36480
67	мен	ес	909	170769	0,00195	0,36675
68	Базаралы	зт	895	171664	0,00192	0,36867
69	бәрі	ес	890	172554	0,00191	0,37058
70	топ	зт	890	173444	0,00191	0,37249
71	соң	шл	885	174329	0,00190	0,37439
72	Әбіш	зт	882	175211	0,00189	0,37628
73	жан	зт	875	176086	0,00188	0,37816
74	жігіт	зт	862	176948	0,00185	0,38001
75	сөйле	ет	856	177804	0,00184	0,38185
76	кез	зт	852	178656	0,00183	0,38368
77	алд	зт	822	179478	0,00177	0,38545
78	қас	зт	813	180291	0,00175	0,38720
79	ой	зт	789	181080	0,00169	0,38889
80	жақсы	сн	766	181846	0,00164	0,39053
81	сияқты	шл	756	182602	0,00162	0,39215
82	бар	ес	754	183356	0,00162	0,39377
83	күл	ет	753	184109	0,00161	0,39538
84	қалың	сн	751	184860	0,00161	0,39699
85	Тәкежан	зт	748	185608	0,00161	0,39860
86	тарт	ет	742	186350	0,00159	0,40019
87	дәл	үс	740	187090	0,00159	0,40178
88	Дәрмен	зт	730	187820	0,00158	0,40336

1	2	3	4	5	6	7
89	ат	ет	729	188549	0,00157	0,40493
90	үст	зт	729	189278	0,00157	0,40650
91	қазір	үс	726	190004	0,00156	0,40806
92	жас	зт	714	190718	0,00153	0,40959
93	жаңағы	сн	694	191412	0,00149	0,41108
94	қазақ	зт	689	192101	0,00148	0,41256
95	қала	зт	689	192790	0,00148	0,41404
96	жібер	ет	684	193474	0,00147	0,41551
97	қара	сн	678	194152	0,00146	0,41697
98	өт	ет	673	194825	0,00145	0,41842
99	жүз	зт	670	195495	0,00144	0,41986
100	ара	зт	669	196164	0,00144	0,42130
101	қыл	ет	668	196832	0,00143	0,42273
102	керек	мд	665	197497	0,00143	0,42416
103	ойла	ет	662	198159	0,00142	0,42558
104	Ербол	зт	660	198879	0,00142	0,42700
105	көңіл	зт	655	199474	0,00141	0,42841
106	аз	мд	652	200126	0,00141	0,42982
107	бет	зт	650	200776	0,00140	0,43122
108	бас	ет	647	201423	0,00139	0,43261
109	үн	зт	642	202065	0,00138	0,43399
110	кір	ет	639	202704	0,00137	0,43536
111	есті	ет	637	203341	0,00137	0,43673
112	әке	зт	633	203974	0,00136	0,43809
113	тап	ет	629	204603	0,00135	0,43944
114	Оразбай	зт	625	205228	0,00134	0,44078
115	қарай	шл	622	205850	0,00134	0,44212
116	өзге	сн	622	206472	0,00134	0,44346
117	мынау	ес	621	207093	0,00133	0,44479
118	барлық	ес	620	207713	0,00133	0,44612
119	халық	зт	614	208327	0,00132	0,44744
120	үшін	шл	608	208935	0,00131	0,44875
121	бүгін	үс	599	209534	0,00129	0,45004
122	Мағаш	зт	596	210130	0,00128	0,45132
123	жұрт	зт	595	210725	0,00128	0,45260
124	тек	шл	589	211314	0,00126	0,45386
125	үш	са	586	211900	0,00126	0,45512
126	түс	ет	572	212472	0,00123	0,45635
127	жалғыз	са	566	213038	0,00122	0,45757
128	ең	үс	564	213602	0,00121	0,45878
129	жатыр	ет	561	214163	0,00120	0,45998
130	пен	шл	560	214723	0,00120	0,46118
131	мал	зт	559	215282	0,00120	0,46238
132	қонақ	зт	558	215840	0,00120	0,46358
133	ұзақ	сн	554	216394	0,00119	0,46477
134	іс	зт	540	216934	0,00116	0,46593
135	аға	зт	539	217473	0,00116	0,46709
136	әлі	үс	538	218011	0,00116	0,46825

1	2	3	4	5	6	7
137	ән	зт	537	218548	0,00115	0,46940
138	мына	ес	537	218085	0,00115	0,47055
139	қыз	зт	533	219618	0,00114	0,47169
140	біз	ес	513	220149	0,00110	0,47279
141	ауыз	зт	513	220662	0,00110	0,47389
142	ақыл	зт	512	221174	0,00110	0,47499
143	көрін	ет	504	221678	0,00108	0,47607
144	соңғы	сн	504	222182	0,00108	0,47715
145	ал	шл	500	222682	0,00107	0,47822
146	қайт	ет	490	223172	0,00105	0,47927
147	болыс	зт	486	223658	0,00104	0,48031
148	Оспан	зт	482	224140	0,00103	0,48134
149	шақ	зт	479	224619	0,00103	0,48237
150	сұра	ет	475	225094	0,00102	0,48441
151	ес	зт	473	225567	0,00102	0,48543
152	бұрын	үс	468	226035	0,00100	0,48643
153	жауап	зт	466	225501	0,00100	0,48743
154	Әйгерім	зт	465	226966	0,00100	0,48843
155	сіз	ес	460	227426	0,00099	0,48942
156	күй	зт	453	227879	0,00097	0,49039
157	оқы	ет	453	228332	0,00097	0,49136
158	осындай	сн	451	228783	0,00097	0,49233
159	кім	ес	440	229223	0,00094	0,49327
160	та	шл	439	229662	0,00094	0,49421
161	тобықты	зт	434	230096	0,00093	0,49514
162	ақ	сн	429	230525	0,00092	0,49606
163	әйел	зт	428	230953	0,00092	0,49698
164	орта	зт	427	231380	0,00092	0,49790
165	таста	ет	426	231806	0,00091	0,49881
166	қайта	үс	421	232227	0,00090	0,49971
167	арт	зт	416	232643	0,00089	0,50060
168	сыр	зт	416	233059	0,00089	0,50149
169	бой	зт	408	233467	0,00088	0,50237
170	жиын	зт	407	233874	0,00087	0,50324
171	әңгіме	зт	406	234280	0,00087	0,50411
172	жыл	зт	401	234681	0,00086	0,50497
173	тіпті	үс	400	235081	0,00086	0,50583
174	жылқы	зт	399	235480	0,00086	0,50669
175	өлең	зт	396	235876	0,00085	0,50754
176	уақыт	зт	392	236268	0,00084	0,50838
177	жақын	сн	388	236656	0,00083	0,51004
178	тіл	зт	387	237043	0,00083	0,51087
179	аса	үс	385	237428	0,00083	0,51170
180	Дәркембай	зт	381	237809	0,00082	0,51252
181	әкел	ет	373	238182	0,00080	0,51332
182	әсіресе	үс	372	238554	0,00080	0,51412
183	сырт	зт	369	238923	0,00079	0,51491
184	ұлық	зт	369	239292	0,00079	0,51570

1	2	3	4	5	6	7
185	қос	ет	363	239655	0,00078	0,51648
186	дос	зт	361	240016	0,00077	0,51725
187	тоқта	ет	359	240375	0,00076	0,51801
188	Бөжей	зт	358	240733	0,00076	0,51877
189	аш	ет	357	241090	0,00076	0,51953
190	талай	үс	357	241447	0,00076	0,52029
191	шеше	зт	357	241804	0,00076	0,52105
192	орын	зт	356	242160	0,00076	0,52181
193	ме	шл	355	242515	0,00076	0,52257
194	мінез	зт	354	242869	0,00076	0,52333
195	ту	ет	354	243223	0,00076	0,52409
196	ұста	ет	354	243577	0,00076	0,52485
197	өл	ет	353	243930	0,00076	0,52561
198	түс	зт	353	244283	0,00076	0,52637
199	түн	зт	352	244635	0,00075	0,52712
200	ендігі	сн	351	244986	0,00075	0,52787
201	шығар	ет	350	245336	0,00075	0,52862
202	қат	ет	348	245684	0,00075	0,52937
203	ас	зт	344	246028	0,00074	0,53011
204	мін	ет	344	246372	0,00074	0,53085
205	кең	сн	339	246711	0,00073	0,53158
206	Баймағамбет	зт	338	247049	0,00072	0,53230
207	анық	сн	336	247385	0,00072	0,53302
208	бай	зт	336	247721	0,00072	0,53374
209	көрсет	ет	334	248055	0,00072	0,53446
210	дес	ет	333	248388	0,00071	0,53517
211	аңғар	ет	330	248718	0,00071	0,53588
212	те	шл	329	249047	0,00071	0,53659
213	мол	сн	328	249375	0,00070	0,53729
214	сонда	үс	328	249703	0,00070	0,53799
215	таны	ет	328	250031	0,00070	0,53869
216	жігітек	зт	327	240358	0,00070	0,53939
217	ба	шл	325	250683	0,00070	0,54009
218	хабар	зт	325	251008	0,00070	0,54079
219	күт	ет	323	251331	0,00069	0,54148
220	он	са	323	251654	0,00069	0,54217
221	соқ	ет	321	251975	0,00069	0,54286
222	Әзімбай	зт	317	252292	0,00068	0,54354
223	шап	ет	316	252608	0,00068	0,54422
224	ти	ет	315	252923	0,00067	0,54489
225	су	зт	313	253236	0,00067	0,54556
226	сәлем	зт	312	253548	0,00067	0,54623
227	істе	ет	312	253860	0,00067	0,54690
228	басқа	сн	309	254169	0,00066	0,54756
229	Шұбар	зт	308	254477	0,00066	0,54822
230	Жиренше	зт	307	254784	0,00066	0,54888
231	өмір	зт	307	255091	0,00066	0,54954
232	жүрек	зт	305	255396	0,00065	0,55019

1	2	3	4	5	6	7
233	орыс	зт	304	255700	0,00065	0,55084
234	жыла	ет	303	256003	0,00065	0,55149
235	шейін	шл	301	256306	0,00065	0,55214
236	жөнел	ет	301	256607	0,00065	0,55279
237	сәт	зт	301	256908	0,00065	0,55344
238	Майбасар	зт	300	257208	0,00064	0,55408
239	жаса	ет	298	257506	0,00064	0,55472
240	үнсіз	сн	298	257804	0,00064	0,55536
241	білдір	ет	297	258101	0,00064	0,55600
242	дала	зт	296	258397	0,00063	0,55663
243	бірге	үс	293	258690	0,00063	0,55726
244	бірі	ес	292	258982	0,00063	0,55789
245	жау	зт	292	259274	0,00063	0,55852
246	қабақ	зт	292	259566	0,00063	0,55915
247	қандай	ес	291	259857	0,00062	0,55978
248	рет	зт	291	260148	0,00062	0,56040
249	тез	үс	290	260438	0,00062	0,56102
250	Тоғжан	зт	290	260728	0,00062	0,56164
251	Ұлжан	зт	290	261018	0,00062	0,56226
252	хат	зт	289	261307	0,00062	0,56288
253	жаңа	үс	286	261593	0,00061	0,56349
254	пе	шл	285	261878	0,00061	0,56471
255	сондай	сн	285	262163	0,00061	0,56532
256	тәрізді	шл	283	262446	0,00061	0,56593
257	қалай	ес	279	262725	0,00060	0,56653
258	есік	зт	278	263003	0,00060	0,56713
259	шай	зт	278	263281	0,00060	0,56773
260	Кәкітай	зт	277	263558	0,00059	0,56832
261	біраз	үс	276	263834	0,00059	0,56891
262	жаз	ет	276	264110	0,00059	0,56950
263	қыс	зт	276	264386	0,00059	0,57009
264	ауыр	сн	273	264659	0,00059	0,57068
265	байлау	зт	268	264927	0,00058	0,57126
266	бері	шл	268	265195	0,00058	0,57184
267	шақыр	ет	266	265461	0,00057	0,57241
268	айнал	ет	265	265726	0,00057	0,57298
269	қон	ет	265	265991	0,00057	0,57355
270	үнде	ет	265	266256	0,00057	0,57412
271	аяқ	зт	264	266520	0,00057	0,57469
272	әуелі	үс	264	266784	0,00057	0,57526
273	же	ет	263	267047	0,00056	0,57582
274	бе	шл	261	267308	0,00056	0,57638
275	Семей	зт	259	267567	0,00056	0,57694
276	қой	зт	257	267824	0,00055	0,57749
277	па	шл	257	268081	0,00055	0,57804
278	әрі	шл	256	268337	0,00055	0,57859
279	ақын	зт	254	268591	0,00055	0,57914
280	сөйлес	ет	254	268845	0,00055	0,57969

1	2	3	4	5	6	7
281	біт	ет	252	269097	0,00054	0,58023
282	кітап	зт	252	269349	0,00054	0,58077
283	от	зт	252	269601	0,00054	0,58131
284	ағайын	зт	251	269852	0,00054	0,58185
285	көк	сн	250	270102	0,00054	0,58239
286	көтере	ет	248	270350	0,00053	0,58292
287	әдейі	үс	245	270595	0,00053	0,58345
288	арнаулы	сн	244	270839	0,00052	0,58397
289	жаңа	сн	244	271083	0,00052	0,58449
290	қош	ет	244	271327	0,00052	0,58501
291	алыс	сн	243	271570	0,00052	0,58553
292	сәл	үс	243	271813	0,00052	0,58605
293	бүгінгі	сн	241	272054	0,00052	0,58657
294	бес	са	240	272294	0,00052	0,58709
295	жатақ	зт	240	272534	0,00052	0,58761
296	ояз	зт	240	272774	0,00052	0,58813
297	жи	ет	239	273013	0,00051	0,58864
298	суық	сн	238	273251	0,00051	0,58915
299	жел	зт	237	273488	0,00051	0,58966
300	ата	ет	236	273724	0,00051	0,59017
301	бойынша	шл	236	273960	0,00051	0,59068
302	дүние	зт	236	274196	0,00051	0,59119
303	қан	зт	236	274432	0,00051	0,59170
304	кеш	зт	233	274665	0,00050	0,59220
305	іш	ет	233	274898	0,00050	0,59270
306	Абылғазы	зт	232	275130	0,00050	0,59320
307	Мәкен	зт	232	275362	0,00050	0,59370
308	жөн	зт	230	275592	0,00049	0,59419
309	сүйсін	ет	230	275822	0,00049	0,59468
310	ұзын	сн	230	276052	0,00049	0,59517
311	кедей	зт	229	276281	0,00049	0,59566
312	туралы	шл	228	276509	0,00049	0,59615
313	байқа	ет	227	276736	0,00049	0,59664
314	қана	шл	227	276963	0,00049	0,59713
315	Павлов	зт	227	277190	0,00049	0,59762
316	алыс	ет	226	277416	0,00049	0,59811
317	қақ	ет	223	277642	0,00048	0,59859
318	Шыңғыс	зт	222	277865	0,00048	0,59907
319	ата	зт	222	278087	0,00048	0,59955
320	қызыл	сн	222	278309	0,00048	0,60003
321	ыстық	сн	222	278531	0,00048	0,60051
322	ие	зт	221	278752	0,00047	0,60098
323	тос	ет	221	278973	0,00047	0,60145
324	ас	ет	220	279193	0,00047	0,60192
325	құдай	зт	220	279413	0,00047	0,60239
326	тең	зт	220	279633	0,00047	0,60286
327	шын	сн	220	279853	0,00047	0,60333
328	артық	сн	219	280072	0,00047	0,60380

1	2	3	4	5	6	7
329	түй	ет	218	280290	0,00047	0,60427
330	шет	зт	218	280508	0,00047	0,60474
331	анау	ес	217	280725	0,00047	0,60521
332	тыңда	ет	217	280942	0,00047	0,60568
333	біреу	ес	216	281158	0,00046	0,60614
334	ерт	ет	216	281374	0,00046	0,60660
335	бұйрық	зт	215	281589	0,00046	0,60706
336	қалып	зт	213	281802	0,00046	0,60752
337	қу	ет	212	282014	0,00046	0,60798
338	Сүйіндік	зт	212	282226	0,00046	0,60844
339	жаман	сн	211	282437	0,00045	0,60889
340	сүй	ет	211	282648	0,00045	0,60934
341	құлақ	зт	209	282857	0,00045	0,60979
342	әмір	зт	208	283065	0,00045	0,61024
343	қадал	ет	208	283273	0,00045	0,61069
344	қыстау	зт	208	283481	0,00045	0,61114
345	асық	ет	207	283688	0,00044	0,61158
346	қамшы	зт	206	283894	0,00044	0,61202
347	Сармолла	зт	206	284100	0,00044	0,61246
348	ер	зт	204	284304	0,00044	0,61290
349	жуан	сн	204	284508	0,00044	0,61334
350	соншалық	үс	204	284710	0,00044	0,61378
351	ұр	ет	202	284912	0,00043	0,61421
352	заман	зт	201	285113	0,00043	0,61464
353	би	зт	200	285313	0,00043	0,61507
354	ұқ	ет	200	285513	0,00043	0,61550
355	бөлек	сн	199	285712	0,00043	0,61593
356	дау	зт	199	285911	0,00043	0,61636
357	алғаш	үс	198	286109	0,00043	0,61679
358	ер	ет	198	286307	0,00043	0,61722
359	жеткіз	ет	198	286505	0,00043	0,61765
360	өзгеше	үс	198	286703	0,00043	0,61808
361	бұрыл	ет	197	286900	0,00042	0,61850
362	кәрі	сн	197	287097	0,00042	0,61892
363	кейде	шл	197	287294	0,00042	0,61934
364	сез	ет	197	287491	0,00042	0,61976
365	айнала	зт	195	287686	0,00042	0,62018
366	айтыл	ет	195	287881	0,00042	0,62060
367	қайда	ес	195	288076	0,00042	0,62102
368	аралас	ет	194	288270	0,00042	0,62144
369	ақырын	үс	193	288463	0,00041	0,62185
370	кейін	үс	193	288656	0,00041	0,62226
371	қар	зт	193	288849	0,00041	0,62267
372	қағаз	зт	192	289041	0,00041	0,62308
373	қай	ес	192	289233	0,00041	0,62349
374	салқын	сн	192	289425	0,00041	0,62390
375	тақа	ет	192	289617	0,00041	0,62431
376	тілек	зт	192	289809	0,00041	0,62472

1	2	3	4	5	6	7
377	ай	зт	191	290000	0,00041	0,62513
378	бақ	ет	191	290191	0,00041	0,62554
379	қосыл	ет	191	290382	0,00041	0,62595
380	төрт	са	190	290572	0,00041	0,62636
381	арыз	зт	189	290761	0,00041	0,62677
382	биік	сн	189	290950	0,00041	0,62718
383	бөкенші	зт	189	291139	0,00041	0,62759
384	жолдас	зт	189	291328	0,00041	0,62800
385	қора	зт	189	291517	0,00041	0,62841
386	құр	ет	189	291706	0,00041	0,62882
387	қоңыр	сн	188	291894	0,00040	0,62922
388	бойы	шл	187	292081	0,00040	0,62962
389	бірнеше	ес	186	292267	0,00040	0,63002
390	Ырғызбай	зт	185	292452	0,00040	0,63042
391	ажар	зт	184	292636	0,00040	0,63082
392	қарсы	сн	183	292819	0,00040	0,63122
393	өнер	зт	183	293002	0,00040	0,63162
394	ізде	ет	183	293185	0,00040	0,63202
395	кішкене	сн	182	293367	0,00040	0,63242
396	түс	зт	182	293549	0,00040	0,63282
397	тыс	зт	182	293731	0,00040	0,63322
398	міне	ес	181	293912	0,00039	0,63361
399	ақ	ет	179	294091	0,00038	0,63399
400	бауыр	зт	179	294270	0,00038	0,63437
401	алғашқы	сн	178	294448	0,00038	0,63475
402	қауым	зт	178	294626	0,00038	0,63513
403	мәлім	сн	178	294804	0,00038	0,63551
404	түгел	үс	177	294981	0,00038	0,63589
405	зор	сн	176	295157	0,00038	0,63627
406	Михайлов	зт	176	295333	0,00038	0,63665
407	еңбек	зт	174	295507	0,00037	0,63702
408	жай	ет	172	295679	0,00037	0,63739
409	қи	ет	172	295851	0,00037	0,63776
410	ертең	үс	171	296022	0,00037	0,63813
411	ет	зт	171	296193	0,00037	0,63850
412	мойын	зт	171	296364	0,00037	0,63887
413	азамат	зт	170	296534	0,00037	0,63924
414	ашық	сн	170	296704	0,00037	0,63961
415	қарсы	үс	170	296874	0,00037	0,63998
416	есеп	зт	169	297403	0,00036	0,64034
417	қысыл	ет	169	297212	0,00036	0,64070
418	ойлан	ет	169	297381	0,00036	0,64106
419	тас	зт	169	297381	0,00036	0,64142
420	қымыз	зт	168	297550	0,00036	0,64178
421	орал	ет	168	297718	0,00036	0,64214
422	ендеше	шл	167	297886	0,00036	0,64250
423	көрші	зт	167	298053	0,00036	0,64286
424	қыр	зт	167	298220	0,00036	0,64322

1	2	3	4	5	6	7
425	дауыс	зт	166	298387	0,00036	0,64358
426	ит	зт	166	298553	0,00036	0,64394
427	төре	зт	166	298719	0,00036	0,64430
428	жайлау	зт	165	298885	0,00035	0,64465
429	кездес	ет	165	299050	0,00035	0,64500
430	немесе	шл	165	299215	0,00035	0,64535
431	салмақ	зт	165	299380	0,00035	0,64570
432	тіле	ет	165	299710	0,00035	0,64605
433	Ысқақ	зт	165	299875	0,00035	0,64640
434	Байдалы	зт	164	300039	0,00035	0,64675
435	Бат	ет	164	300203	0,00035	0,64710
436	Ділдә	зт	164	300367	0,00035	0,64745
437	аңда	ет	163	300530	0,00035	0,64780
438	қаш	ет	163	300693	0,00035	0,64815
439	өлім	зт	163	300856	0,00035	0,64850
440	сайын	шл	163	301019	0,00035	0,64885
441	сонымен	шл	163	301182	0,00035	0,64920
442	тау	зт	163	301345	0,00035	0,64955
443	жиыл	ет	162	301507	0,00035	0,64990
444	осылай	үс	162	301669	0,00035	0,65025
445	сары	сн	162	301831	0,00035	0,65060
446	төсек	зт	162	301993	0,00035	0,65095
447	е	од	161	302154	0,00035	0,65130
448	ар	сн	160	302314	0,00035	0,65165
449	ана	зт	159	302473	0,00034	0,65199
450	белгі	зт	159	302632	0,00034	0,65233
451	байла	ет	158	302790	0,00034	0,65267
452	жұмыс	зт	158	302948	0,00034	0,65301
453	қой	шл	158	303106	0,00034	0,65335
454	қоныс	зт	158	303264	0,00034	0,65369
455	қоста	ет	158	303422	0,00034	0,65403
456	киім	зт	157	303579	0,00034	0,65437
457	патша	зт	157	303736	0,00034	0,65471
458	шапшаң	үс	157	303893	0,00034	0,65505
459	ағай	зт	156	304049	0,00033	0,65538
460	бел	зт	156	304205	0,00033	0,65571
461	кейде	үс	156	304361	0,00033	0,65604
462	қазіргі	сн	156	304517	0,00033	0,65637
463	аппақ	сн	155	304672	0,00033	0,65673
464	таң	зт	155	304827	0,00033	0,65706
465	тара	ет	155	304982	0,00033	0,65739
466	денелі	сн	154	305136	0,00033	0,65772
467	жар	зт	154	305290	0,00033	0,65805
468	кеуде	зт	153	305443	0,00033	0,65838
469	күлкі	зт	153	305596	0,00033	0,65871
470	ұш	ет	153	305749	0,00033	0,65904
471	бөлме	зт	152	305901	0,00033	0,65937
472	ашыл	ет	151	306052	0,00032	0,65969

1	2	3	4	5	6	7
473	мұң	зт	151	306203	0,00032	0,66001
474	атаулы	үс	150	306353	0,00032	0,66033
475	жандарал	зт	150	306503	0,00032	0,66065
476	жүз	са	150	306653	0,00032	0,66097
477	қайыр	ет	150	306803	0,00032	0,66129
478	айтыс	ет	149	306952	0,00032	0,66161
479	ұры	зт	149	307101	0,00032	0,66193
480	арман	зт	148	307249	0,00032	0,66225
481	таныт	ет	148	307397	0,00032	0,66257
482	жаз	зт	147	307544	0,00032	0,66289
483	солай	үс	147	307691	0,00032	0,66321
484	уақ	зт	147	307838	0,00032	0,66353
485	бұр	ет	146	307984	0,00031	0,66384
486	Мағыш	зт	146	308130	0,00031	0,66415
487	қараңғы	сн	144	308274	0,00031	0,66446
488	амандас	ет	143	308417	0,00031	0,66477
489	ауыс	ет	143	308560	0,00031	0,66508
490	жаға	зт	143	308703	0,00031	0,66539
491	сақал	зт	143	308846	0,00031	0,66570
492	ауру	зт	142	308988	0,00030	0,66600
493	әже	зт	142	309130	0,00030	0,66630
494	кезек	зт	142	309272	0,00030	0,66660
495	еркек	зт	140	309412	0,00030	0,66690
496	тоқтат	ет	140	309552	0,00030	0,66720
497	шеш	ет	139	309691	0,00030	0,66750
498	көпшілік	зт	138	309829	0,00030	0,66780
499	үнемі	үс	138	309967	0,00030	0,66810
500	іні	зт	138	310105	0,00030	0,66840

ӘДЕБИЕТ

1. Абай тілі сөздігі. Алматы, 1968. 734-б.
2. *Алексеев П. М.* Распределение лексических единиц по длине в тексте и словаре // Квантитативная лингвистика и автоматический анализ текстов. Тарту, 1986. Вып. 745. С. 3-28.
3. *Ахабаев А.* Статистический анализ лексико-морфологической структуры языка казахской публицистики: Автореф. дисс. канд. филол. наук. Алма-Ата, 1971. С. 24.
4. *Ахметов Е.Г.* Статистическая характеристика родительного падежа имени существительного в английском и казахском языках // Тезисы докладов и сообщений Всесоюзного семинара "Статистическое и информационное изучение тюркских языков". Алма-Ата, 1969. С. 51-54.
5. *Байтанаева Д.А., Бектаев К.Б.* Энтропия казахского текста // Статистика казахского текста. Алма-Ата, 1973. С. 664-696.
6. *Байтанаева Д.А.* Энтропия распределения слогов казахского письменного текста // Тезисы докладов и сообщений Всесоюзного семинара "Статистическое и информационное изучение тюркских языков". Алма-Ата, 1969. С. 80-82.
7. *Байтанаева Д.А.* Информационные характеристики казахского текста. Алма-Ата, 1985. С. 18.
8. *Балакаев М.* Современный казахский язык. Алма-Ата, 1985.
9. *Бектаев К.Б.* Алфавитно-частотный словарь слогов казахского языка // Статистика казахского текста. Алма-Ата, 1973. С. 566-612.
10. *Бектаев Қ., Белботаев А., Смайылов Т.* Қазақ тіл біліміндегі лингвистиканың алғашқы көріністері // Құдайберген Жұбанов және қазақ совет тіл білімі. Алматы: Ғылым, 1990. 108-115-бб.
11. *Бектаев К.Б., Пиотровский Р.Г.* Математические методы в языкознании. Ч.1. Теория вероятностей и моделирование нормы языка. Алма-Ата, 1973. 281 с.; Математическая статистика и моделирование текста. Алма-Ата, 1974. 334 с. Ч. II. Алма-Ата, 1974.
12. *Бектаев К.Б.* Статистико-информационная типология тюркского текста. Алма-Ата, 1978. 183 с.
13. *Бектаев Қ.Б., Жұбанов А.Қ., Мырзабеков С., Белботаев А.Б.* М.Әуезовтің "Абай жолы" романының жиілік сөздігі. Алматы, 1979. 334-б.
14. *Бектаев Қ.Б., Жұбанов А.Қ., Мырзабеков С., Белботаев А.Б.* М.О.Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздіктері. Алматы-Түркістан, 1995. 346-б.
15. *Бектаев К.Б., Зубов А.В., Ковалевич Е.Ф.* и др. К исследованию законов распределения лингвистических единиц // Статистика текста. Минск, 1969. С. 131-162.
16. *Бектаев К.Б., Лукьяненков К.Ф.* О законах распределения единиц письменной речи // Статистика речи и автоматический анализ текста. Л., 1971. С. 47-112.
17. *Бектаев Қ.* Ықтималдықтар теориясы және математикалық статистика. Алматы: Рауан, 1991. 432 с.
18. *Белботаев А.Б.* Лингвостатистические характеристики частей речи казахского текста: Автореф. дисс. ... канд. филол. наук. Алма-Ата, 1992. С. 32.
19. *Белботаев А.Б.* Ғылыми-техникалық стильдегі сын есім аффикстерінің қолданылуы туралы. 225-230-бб.; Математика тексінің алфавитті-жиілік сөздігі. 543-552-бб. // Қазақ тексінің статистикасы. Алматы: Ғылым, 1973.
20. *Белоногов Г.Г.* О некоторых статистических закономерностях в русской письменной речи // ВЯ. 1962. №1. С. 100-101.
21. *Беляева Л.Н.* Применение ЭВМ в лингвистических исследованиях и лингводидактике. Л., 1986.
22. *Бодуэн де Куртенэ И.А.* Языкознание или лингвистика XIX века // Избранные труды по общему языкознанию. М., 1963. Т.2.
23. *Вейсалов Ф. Е., Керимов Я.К., Сафаров М.Г.* Комбинаторика фонем современного азербайджанского языка // Вопросы азербайджанского языкознания. Баку, 1981.
24. *Вентцель Е.С.* Теория вероятностей. М., 1969. 576 с.
25. *Винер Н.* Кибернетика. М., 1983.

26. *Виноградов В.В.* Современный русский язык. Введение в грамматическое учение о слове. 1938. Вып. 1. 160 с.
27. *Гарипов Т. М.* Лексикостатистические отношения тюркских языков Урало-Поволжья // Языковые контакты в Башкирии. Уфа, 1972. С. 248-267.
28. *Герд А.С.* Предмет и основные направления прикладной лингвистики // Прикладное языкознание. С.-Пб., 1996. С. 5-14.
29. *Гмурман В.Е.* Теория вероятностей и математическая статистика. М.: Высшая школа, 1972. 368 с.
30. *Головин Б.Н.* Язык и статистика. М., 1971. 191 с.
31. *Джубанов А.Х.* Квантитативная структура казахского текста (опыт лингвистического анализа на ЭВМ). Алма-Ата, 1987. С. 147.
32. *Джубанов А.Х.* К вопросу о графемной статистике казахского текста // Вопросы казахской фонетики и фонологии. Алма-Ата, 1979. С. 79-86.
33. *Джубанов А.Х., Джунисбеков А.* Преобразование с помощью ЭВМ казахских текстов в фонетическую запись // Материалы семинара "Статистическая оптимизация преподавания языков и инженерная лингвистика". Чимкент, 1980. С. 317-319.
34. *Джунусбеков А.* Гласные казахского языка. Алма-Ата, 1972. С. 94.
35. *Длин А. М.* Математическая статистика в технике. М., 1958. 466 с.
36. *Ергалауов А.* Статистическая характеристика форм страдательного залога в английском и казахском языках // Тезисы докладов и сообщений Всесоюзного семинара "Статистическое и информационное изучение тюркских языков". Алма-Ата, 1969. С. 62-63.
37. *Estoup J.* Gammes stenographiques. Paris, 1916.
38. *Жетешиков Ж.* Статистический анализ словоизменительных аффиксов имен существительных киргизского языка (на материале газетных текстов)//Материалы семинара "Статистическая оптимизация преподавания языков и инженерная лингвистика". Чимкент, 1980. С. 25.
39. *Жубанов А.К.* Основные принципы формализации содержания казахского текста. Алматы, 2002. 250 с.
40. *Жұбанов Қ.* Қазақ тілі жөніндегі зерттеулер. Алматы: Ғылым,1999. 581 б.
41. *Засорина Л.Н.* Частотные словари и вопросы лексико-статистики // Межвузовская конференция по вопросам частотных словарей и автоматизации. Л., 1966. С. 3-4.
42. *Звегинцев В.А.* 1) История языкознания XIX–XX веков о очерках и извлечениях. Ч. II. М., 1965; 2) Теорическая и прикладная лингвистика. М., 1968.
43. *Зекенова А.* М.О. Әуезов пьесаларындағы есім сөздердің морфологиялық құрылысы жайында // Известия АН КазССР. Сер. филолог. 1976. С. 56-62.
44. Знание языка и языкознание. М., 1991.
45. *Зубов А.В.* Основы лингвистической информатики: Учеб. пособие. Ч.I. Минск, 1991; Ч. II. 1992; Ч. III. 1993.
46. *Зубов А.В., Лихтарович А.А.* ЭВМ анализирует текст. Минск, 1989.
47. *Зубов А.В., Хотяшов Э.Н.* Статистический анализ текста с помощью ЭВМ // Энтропия языка и статистика речи. Минск, 1966. С. 118-166.
48. *Ибрагимов С.И.* Некоторые статистические характеристики имен существительных киргизского языка // Тезисы докладов и сообщений всесоюзного семинара "Статистическое и информационное изучение тюркских языков". Алма-Ата, 1969. С. 34-38.
49. *Ибрагимов Т.И.* Некоторые статистические данные о слогах татарского языка: вероятностные методы и кибернетика // Учен. зап. Казанского ун-та. Вып. 4. Т. 25. Кн. 6. 1966. С. 74-78; *Его же.* Изучения слогов и структуры их сочетаний в татарском литературном языке. Казань, 1970.
50. Использование ЭВМ в лингвистических исследованиях / Под ред. В.И.Перебейнос. Киев, 1990.
51. *Ишанов К.И., Садиков А.В., Мырзабеков С.* Частотно-сопоставительная характеристика падежных форм русского и казахского языков // Тезисы докладов и сообщений Всесоюзного семинара "Статистическое и информационное изучение тюркских языков". Алма-Ата, 1969. С. 56-58.
52. *Кеңесбаев І., Мұсабаев Ғ.* Қазіргі қазақ тілі. Алматы, 1962. 315-б.

53. *Кенесбаева С.С.* К вопросу исследования протетического звука в словах арабского происхождения статистическим методом // Тезисы докладов и сообщений Всесоюзного семинара "Статистическое и информационное изучение тюркских языков". Алма-Ата, 1969. С.43-44.
54. *Клоусон ДЖ.* Лексико-статистическая оценка алтайской теории // ВЯ. 1969. №5. С.22-41.
55. Компьютеризация общества и человеческий фактор / Под ред. *А.И. Ракитова. М., 1988.*
56. *Кондратов А.М.* Электронный разум. М., 1987.
57. *Котов Р.Г., Якушин Б.В.* Языки информационных систем. М., 1979.
58. Қазақ тексінің статистикасы. Алматы, 1973. 731-б.
59. Қазақ тілінің түсіндірме сөздігі. 1-10-томдар. Алматы: Ғылым, 1974–1986 жж.
60. *Қалыбеков Б.Е.* 50-жылдардағы бастауыш сынып оқулықтарының лексикалық жүйесі мен морфологиялық құрылымының статистикасы: Афтореф. дисс. ...канд. наук. Алматы, 2003. 25 с.
61. *Құрышжанов А.Қ., Жұбанов А.Қ., Белботаев А.Б.* Куманша-қазақша жиілік сөздік. Алматы: Ғылым, 1978. 277 б.
62. Лингвистический энциклопедический словарь. М., 1990.
63. *Лукьяненков К.Ф.* Использование схем Пуассона и Гаусса в исследовании распределения лингвистических единиц текста // Вопросы лингвостатистики и автоматизация лингвистических работ. М., 1970. Вып. 2. С. 3-14.
64. *Маматов Д.* Количественный анализ сложных согласных // Опыт экспериментального и структурного изучения языка. Ташкент, 1982. С. 108-111.
65. *Мандельброт Б.* О рекуррентном кодировании, ограничивающем влияние помех// Теория передачи сообщений. М., 1957. С. 139-157.
66. *Марчук Ю.Н.* Математические методы в языкознании. М., 1990.
67. *Молдабеков К.* Лингвостатистические исследования казахских текстов для младших школьников (на материале текстов учебников начальных классов и литературы для детей): Дисс. канд. филол. наук. Алма-Ата, 1985, 259 с.
68. *Мұсабаев Ғ.* Қазіргі қазақ тіліндегі сын есімнің шырайлары. Алматы, 1951. 90 б.
69. *Мырзабеков С.* Қазақ тілін зерттеуде санды пайдалану // Қазақ тілі грамматикасы бойынша зерттеулер. Алматы, 1975.
70. *Мырзабеков С.* Статистико-лингвистический анализ структуры глагола современного казахского языка: Автореф. дисс. ... канд. филол. наук. Алма-Ата, 1973. С. 32.
71. *Налимов В.В.* Вероятностная модель языка. М., 1974.
72. *Насыров Д.Х., Аймбетов А.* О статистике фонем каракалпакского языка // Материалы семинара "Статистическая оптимизация преподавания языков и инженерная лингвистика". Чимкент, 1980. С. 273-274.
73. Новое в зарубежной лингвистике. Вып. XXIV. Компьютерная лингвистика. М., 1989.
74. *Пиотровская А.А.* Автоматическое приведение именных словоупотреблений к канонической форме // НТИ. Сер. 2. 1977. №1. С.32-36.
75. *Пиотровский Р.Г.* Экстралингвистические и внутриязыковые вопросы при переработке текста в системе "человек-машина-человек"// Вопросы социальной лингвистики. Л., 1969. С.40-64.
76. *Пиотровский Р.Г.* Информационные измерения языка. Л., 1969. 116 с.
77. *Пиотровский Р.Г.* 1) Текст, машина, человек. Л., 1978; 2) Инженерная лингвистика и теория языка. Л., 1979.
78. Прикладные аспекты лингвистики. М., 1989.
79. *Ризаев С.А.* Из опыта исследования статистической структуры слога в узбекском языке // Общественные науки в Узбекистане. Ташкент, 1972. №1; *Его же.* Статолингвистический анализ структуры слова в современном узбекском языке // Материалы семинара "Статистическая оптимизация преподавания языков и инженерная лингвистика". Алма-Ата, 1980. С. 180-182.
80. *Романовский В.И.* Математическая статистика. Ташкент, 1961. Кн. 1. 637 с.; 1963. Кн. 2. 794 с.
81. *Романовский В.И.* Применение математической статистики в опытном деле. М.-Л., 1947. 247с.
82. *Рогожникова Р.П., Чернышева Л.В., Кузнецова Е.* Основные направления автоматизация лингвистических исследований. Л., 1988.

83. *Рождественский Ю.В.* Техника, культура, язык. М., 1993.
84. *Садыков Т.* Проблемы моделирования тюркской морфологии (аспект порождения киргизской именной словоформы). Фрунзе, 1987. С. 121. Стуктурная и прикладная лингвистика. Вып. / Под ред. А.С.Герда. Л., 1978.
85. *Смирнов Н.В., Дунин-Барковский И.В.* Курс теории вероятностной и математической статистики. М., 1969. 511с.
86. *Струве П.Б.* Кто первый указал на применение статистики к филологическим исследованиям? // Изв. Российской АН. Сер. 6. 1918. №13.
87. *Татубаев С.С.* Статистические подходы к исследованию певческой фонетики казахского языка // Тезисы докладов и сообщений Всесоюзного семинара "Статистическое и информационное изучение тюркских языков". Алма-Ата, 1969. С. 38-40.
88. *Феллер В.* Введение в теорию вероятностей и ее приложения. М., 1967. Т.1. 498 с.
89. *Филин Ф.П.* О некоторых философских вопросах языкознания // Ленинизм и теоретические проблемы языкознания. М., 1970.
90. *Фрумкина Р.М.* Статистические методы изучения лексики. М.,1964. 115 с.
91. *Фрумкина Р.М.* Роль статистических методов в современных лингвистических исследованиях // Математическая лингвистика. М., 1973. С. 156 и след.
92. *Фрумкина Р.М.* Вероятность элементов текста и речевое поведение. М., 1971. 168 с.
93. *Фрумкина Р.М.* О законах распределения слов и классов слов // Структурно-типологические исследования. М., 1962. С. 124-133.
94. *Хальд Л.* Математическая статистика с техническим приложением. М., 1956. 644 с.
95. *Шингарева Е.А.* Семиотические основы лингвистической информатики. Л., 1987.
96. *Шор Я.Б.* Статистические методы анализа и контроля качества и надежности. М., 1962. 552 с.
97. *Штиндлова И.* Обратные словари // Автоматизация в лингвистике. М.-Л., 1966. 87-88-бб.
98. *Ысқақов А.* Қазіргі қазақ тілі. Алматы: Ана тілі, 1991. 384 б.
99. Энтропия языка и статистика речи. Минск, 1966.
100. Словарь языка Пушкина. М., 1956-1961.Т. 1-4.
101. *Генкель М.А.* Частотный словарь романа Д.Н. Мамина-Сибиряка "Приваловские миллионы". Пермь, 1974.

ОСНОВНЫЕ ПРИНЦИПЫ ФОРМАЛИЗАЦИИ СОДЕРЖАНИЯ КАЗАХСКОГО ТЕКСТА



Книга посвящена новой области языкознания – лингвистике текста. Рассмотрены такие проблемные вопросы казахского языкознания, как текст и его конструктивные элементы с точки зрения формальной и содержательной структуры его формирования. Исследование преследует цель – получение семантического представления казахского текста и установление основных формальных принципов его генерации, т.е. решение вопросов построения и/или порождения казахского текста.

Монография рассчитана на научных работников в области структурной, прикладной и математической лингвистики, а также преподавателей, студентов и аспирантов филологических и математических специальностей высших учебных заведений.

• ПРЕДИСЛОВИЕ

• *Глава первая*
ТЕКСТ И ЕГО ОСНОВНЫЕ ЗНАКОВЫЕ ХАРАКТЕРИСТИКИ

• *Глава вторая*
ФОРМАЛЬНОЕ ПРЕДСТАВЛЕНИЕ СЛОВАРЯ КАЗАХСКОГО ЯЗЫКА

• *Глава третья*
СЕМАНТИЧЕСКАЯ КЛАССИФИКАЦИЯ ЗНАМЕНАТЕЛЬНЫХ КЛАССОВ СЛОВ КАЗАХСКОГО ЯЗЫКА

• *Глава четвертая*
АБЗАЦ КАК ОСНОВНОЙ СЕМАНТИКО-СИНТАКСИЧЕСКИЙ КОМПОНЕНТ КАЗАХСКОГО ТЕКСТА

• *Глава пятая*
ОСНОВНОЕ СТАТИЧЕСКОЕ СОДЕРЖАНИЕ ТЕКСТА И ЕГО ФОРМАЛЬНОЕ ОПИСАНИЕ

• ЗАКЛЮЧЕНИЕ

• ПРИЛОЖЕНИЯ

• ЛИТЕРАТУРА

ПРЕДИСЛОВИЕ

С провозглашением казахского языка как государственного возрастает его роль в управлении народным хозяйством, обучении и воспитании, культурной жизни и информационном обеспечении суверенной республики.

Являясь высокоразвитой республикой, Казахстан не может остаться в стороне от нового этапа научно-технической революции, связанного с широким внедрением во все аспекты жизни республики компьютерной техники. Если в промышленности такая техника используется достаточно успешно, то информационные технологии на базе казахского языка применяются еще очень слабо.

Эффективное развитие всех областей науки в республике невозможно без создания автоматизированных систем управления народным хозяйством. Основой таких систем в новых условиях должен стать казахский язык. Широкому распространению казахской художественной и научной литературы будет в значительной степени способствовать создание автоматизированных поисковых систем в библиотеках, на предприятиях, в вузах, школах. Естественно, что такие системы должны функционировать на базе казахского языка. Значительно может ускорить процесс активного внедрения в практику казахского языка создание автоматизированных систем обучения в школах и вузах. Как новое средство обучения компьютер вызывает большой интерес, дает значительный эффект при обучении.

Одной из основных причин недостаточного использования компьютеров в информационных технологиях на базе казахского языка и в его обучении является отсутствие формальных описаний казахского языка, которые могли бы быть восприняты электронной машиной. Существующие грамматики казахского языка создавались в свое время для их использования человеком. Правила таких грамматик не формализуемы или плохо формализуемы, так как в основном обращены к интуиции носителя языка, его лингвистическим и экстралингвистическим знаниям. Такими свойствами современные компьютеры не обладают, поэтому необходимо создавать описания языков, в том числе и казахского, которые были бы им понятны.

Создание подобных описаний наталкивается на ряд особенностей, специфичных для казахского языка. К таким особенностям относятся сложность, неоднокодность, недетерминированность и нечеткость языка [143, с.3-4]. Сложность казахского языка заключается в наличии в нем большого числа разнородных единиц (морфем, слов и т.д.), способных вступать в огромное число разнообразных отношений. Суть неоднозначности языка в том, что при любом способе классификации этих единиц число таких классов достаточно велико и разброс количества единиц, входящих в эти классы, также достаточно широк. Недетерминированность казахского языка здесь выражается в том, что отношения между формой единицы и ее

содержанием (словом и его значением, предложением и его смыслом, текстом и его содержанием) многозначны. Наконец, недетерминированность казахского языка заключается в нечеткости областей значений слов и выражений, в нечеткости границ синтаксической правильности выражений этого языка.

Эти особенности казахского языка и создают значительные трудности при построении формальных описаний для его использования в системах автоматического анализа и синтеза казахского текста, а также в автоматизированных системах обучения языку и гуманитарным дисциплинам на базе казахского языка.

В плане традиционного метода исследования грамматического строя казахского языка ученые не ставили перед собой цель применения его в автоматизированных системах, поэтому непосредственное использование в нашей работе имеющихся фактов не всегда представляется возможным. Все же некоторые данные, полученные казахстанскими учеными, брались нами за основу.

Казахстанские ученые при изучении грамматического строя казахского языка сталкиваются с различными проблемами. Следует отметить, что подобное изучение казахского языка, хотя в той или иной мере затрагивалось в первых учебниках и отдельных публикациях, только с 30–40-х годов оно стало приобретать научно-исследовательский характер. В эти годы вышла в свет работа Н. Т. Сауранбаева [249], посвященная сложным предложениям. Некоторые вопросы сложного предложения, а именно классификации придаточных предложений, были проанализированы С.Жиенбаевым [104].

В синтаксической системе любого языка особое место занимает структура простого предложения, ибо общие закономерности, присущие синтаксису, в основном вытекают из природы простого предложения. Эта область синтаксиса была подробно изучена М. Балакаевым. Благодаря его исследованиям, в казахском языкознании появился совершенно новый труд, посвященный синтаксису словосочетаний казахского языка. В этом и в других своих исследованиях ученый сумел доказать, что типы словосочетаний представляют собой особую отрасль синтаксиса казахского языка [24,25].

Исследователи казахского языка впоследствии стали выдвигать много интересных проблем, связанных со словосочетаниями и предложениями. В последние годы различные виды предложений исследовались К. Есеновым, согласно данным которого, в предложении конструируются такие главные закономерности, присущие синтаксису, как сочетаемость слов, синтагматические группы, синтаксические группы, синтаксические обороты, связь с другими предложениями и т.д. [102, 103].

Ученые в области синтаксиса тюркских языков, изучая актуальные проблемы сложных предложений, стали предлагать включить в парадигму сложных предложений также текст, в составе которого обычно рассматривается и сложное синтаксическое целое, и абзац, и текст с прямой и косвенной речью [115].

Текст – сложная единица, содержащая основные единицы языка: морфемы, слова, словосочетания, предложения. В связи с этим, невозможно построить модель, позволяющую компьютеру понять содержание текста без разработки правил формализации значений этих основных языковых единиц.

Цель нашего исследования - поиск основных принципов формализации содержания казахского текста как общепринятой единицы общения людей. Такая постановка вопроса относительно казахского языка ставится на повестку дня

впервые. Объектом данного исследования являются казахские письменные тексты. В их число входят:

1) рассказы из художественных текстов известных казахских писателей [166, с.3-154]; они охватывают 20 минитекстов, включающих 500 абзацев;

2) тексты научно-популярного стиля [148, 167]; общее число абзацев этих текстов составило также 500;

3) публицистические тексты [68]; в сумме исследовано 35 газетных статей – 502 абзаца. Все газетные материалы одной тематики, идущие под рубрикой "Алты жасарлар мектепке барады" ("Шестилетние идут в школу") и касающиеся новой школьной программы.

Таким образом, по текстам трех стилей рассматривались структуры около 1500 абзацев. Выбор именно этих текстов объясняется тем, что они представляют три наиболее употребительных функционально-смысловых типа речи (повествование, рассуждение, описание), без детального изучения которых невозможно построить универсальную модель казахского текста, позволяющую распознать его содержание. Исходя из этого, в работе в рамках единой теории текстообразования последовательно рассматриваются проблемы лингвистики текста как основного вопроса современной лингвистики. Решаются задачи формального представления словаря казахского языка и абзаца как основной семантически значимой текстовой единицы. В главе, в которой приведена семантическая классификация знаменательных классов слов казахского языка, по существу, представлена модель лексикона человека, являющегося основой для порождения любого казахского текста. В последующих главах представлены результаты классификационного и статистического исследования конкретных казахских текстов по их функционально-смысловому и предметно-логическому содержанию в структуре абзацев.

В заключительных главах книги речь идет о статическом содержании текста, т.е. о, так называемых, опорных словах текста и способах их определения, поскольку вокруг этих слов сконцентрировано основное статическое содержание текста. Кроме того, в последней главе даны некоторые возможные случаи наиболее вероятных формул казахского текста и вводные понятия о семантико-синтаксическом языке СЕМСИНТ. Здесь, в основном, речь идет о некоторых теоретических сторонах этого вопроса, практическая реализация которых будет иметь продолжение.

Итак, исходной идеей исследования явилось положение о том, что в процессе речевой деятельности человека происходит сложное переплетение детерминированных и вероятностных процессов как на лексическом, так и на синтаксическом уровнях. Этот факт связан с социальным характером языка и его постоянным развитием [252, с.118]. Задача нашего исследования состояла в выделении тех семантически значимых единиц казахского текста и правил их комбинирования, которые могли быть основой детерминированной (алгоритмической) информации и тех единиц и правил, которые можно варьировать по определенным вероятностным законам. Полагаем, что рассмотренные в работе основные принципы формализации содержания казахского текста и другие материалы исследования помогут решению задач по автоматическому анализу и синтезу казахских текстов в информационных, исследовательских и учебных целях.

Глава первая
ТЕКСТ И ЕГО ОСНОВНЫЕ ЗНАКОВЫЕ ХАРАКТЕРИСТИКИ

1.1. Язык, речь
Основные функции и единицы

Общепризнанно, что естественный язык по существу есть не что иное, как инструмент, позволяющий индивидуму строить во внешнем мире знаковые модели, более или менее адекватно объективирующие фрагменты его концептуальной системы.

Принципы рассмотрения естественного языка как орудия, инструмента, средства общения не новы для языкознания. По этому поводу еще высказывался В. Гумбольдт: "Язык происходит вследствие внутренней потребности человека, а не внешней только надобности" [89, с.10]. Но все же распространение идеи о том, что язык представляет собой функциональную материальную систему семиотического, или знакового характера, а также вопросы использования языка как средства общения связывают в первую очередь с Ф. де Соссюром и его "Курсом общей лингвистики". Только одно из основных положений его идей о том, что "язык есть форма, а не субстанция" [259, с.120], предопределило судьбу ряда направлений языкознания на много лет вперед. Эти разделы языкознания были объединены впоследствии под общим названием "структурализм". Общеизвестно и то, что разграничение понятий "язык" и "речь" впервые было выдвинуто и обосновано Ф. де Соссюром [259, с.33-43] и в дальнейшем развито другими учеными, в частности советским лингвистом Л. В. Щербой [323, с.361-373; 324, гл. 6]. "Разделяя язык и речь, – пишет Ф. де Соссюр, – мы тем самым отделяем: 1) социальное от индивидуального; 2) существенное от побочного и более или менее случайного" [260, с.52]. Это говорит о том, что Ф. де Соссюр на первое место в характеристике языка ставил его социальный характер. Оценивая данный факт, Н. А. Слюсарева замечает: "Объяснение социального было новым в лингвистике, и в этом Соссюр опередил современное языкознание на полвека" [258, с.13].

Определение языка как "важнейшего средства человеческого общения" относится к более общим понятиям. Поэтому, учитывая сложность образования языка, В. М. Солнцев конкретизирует определение в зависимости от того, с какой точки зрения оно может быть рассмотрено, т. е. какую именно сторону (или стороны) языка требуется выделить. В. М. Солнцев в книге "Язык как системно-структурное образование" в качестве примера приводит пять таких определений языка, рассматривая их с различных точек зрения. На наш взгляд, эти определения отражают многие характеристики языка, которыми мы будем не раз пользоваться в дальнейших изложениях, поэтому процитируем их. В. М. Солнцев пишет: "... а) с точки зрения функции языка: язык есть средство общения людей и, как таковое, есть средство формирования, выражения и сообщения мысли; б) с точки

зрения устройства (механизма) языка: язык есть набор некоторых единиц и правил использования этих единиц, т. е. комбинирования единиц; в) с точки зрения существования языка: язык есть результат социального, коллективного навыка "деления" единиц из звуковой материи путем сопряжения некоторых звуков с некоторым смыслом; г) с семиотической точки зрения: язык есть система знаков, т. е. система материальных предметов (звуков), наделенных свойством обозначать что-то, существующее вне их самих; д) с точки зрения теории информации: язык есть код, с помощью которого кодируется семантическая информация, и др." [261, с.11-12].

По поводу разграничения понятия "язык" и "речь" В. А. Звегинцев полагал: "Такое разграничение истолковывается учеными по-разному, но, несмотря на это, оно должно носить абсолютный характер" [119, с.235]. Указывая на различия между языком и речью на уровне предложения, он отмечал целенаправленность и ситуативность речи. Кроме того, он считал не правомерным противопоставление языка и речи по принципу их ограниченности и неограниченности. По его мнению, данный принцип, наоборот, свидетельствует о сближении языка и речи [119, с.235]. Подчеркивая различные признаки, сближающие язык и речь, он писал: "...хотя лингвистика изучает "две вещи", два разных явления, они взаимосвязаны и взаимозависимы, а поэтому объект у нее один, но двойственной природы" [119, с.240]. И далее, продолжая указанную мысль, он сделал вывод о том, что проблему взаимосвязи языка и речи, а также проблему взаимосвязи языка и мышления необходимо решать с учетом всего взаимодействующего триединства, т.е. языка, речи и мысли [119, с.249]. Следует отметить, что язык выступает как орудие непосредственного формирования мысли, и потому языковое общение лежит в основе других форм и видов информации. Среди всех знаковых систем именно язык является для человека самым естественным, самым привычным "вместилищем мысли". Как писали К. Маркс и Ф. Энгельс: "Язык есть непосредственная действительность мысли" [188, с.448]. "Язык, будучи средством коммуникации и одновременно средством реализации мысли, – отмечал И. Р. Гальперин, – должен в своем статическом и динамическом проявлении, в своих формах и в их применении отражать закономерности мыслительного процесса, который, в свою очередь, отражает явления объективной действительности" [70, с.20].

Язык существует в силу того, что он функционирует, а именно в речи, в высказываниях и в речевом акте. В связи с этим, Ю. С. Маслов полагал: "Речь есть форма существования языка. Язык функционирует и "непосредственно дан" в речи. Но в отличие от речи он есть абстрактная сущность" [189, с.9].

Язык, как известно, есть система элементов, находящаяся в распоряжении того или иного коллектива, т.е. единицы разных ярусов вместе с системой правил функционирования этих единиц (грамматика этого языка). Подчеркивая социальный характер языка, Ф. де Соссюр писал: "Язык есть нечто социальное, по существу независимое от индивида", – и далее, продолжая эту мысль он отмечал, что язык "существует в коллективе как совокупность отпечатков, имеющихся у каждого в голове, наподобие словаря, экземпляры которого вполне тождественные, находились бы в пользовании многих лиц" [260, с.57], "язык существует только в силу своего рода договора, заключенного членами коллектива" [260, с.52-53].

Устанавливая определенные границы между функциями языка и речи, А.А.Леонтьев предлагал понимать под функциями языка "лишь те функциональные характеристики речевой деятельности, которые проявляются в любой речевой

ситуации. Эти функции, как правило, не имеют в языке соответствующих им закрепленных за ними элементов..." [174, с.101; 175, с.32].

В модели речевого акта Р. Якобсона языку приписываются следующие функции: эмотивная, конативная, референтная, метаязыковая, фатическая. По А. А. Леонтьеву, в данной модели в одном ряду рассматриваются такие обычно различные функции языка, как коммуникативная и эстетическая. К функциям языка он относит коммуникативную функцию, функцию интеллективной деятельности, функцию овладения общественно-историческим опытом человечества, национально-культурную и функцию познания [174, с.106]. Под "коммуникативной функцией" здесь необходимо понимать "предназначенность языковых средств для удовлетворения потребностей в обмене языковой информацией (а не односторонней передаче сообщений)" [146, с.39].

Основываясь на указанных А. А. Леонтьевым признаках функций языка, Л. А. Киселева установила иерархическую систему функций языка [146, с.40-44]. По ее мнению, мыслительная и коммуникативная функции языка образуют двуединство, "детерминированное диалектическим единством потребностей мыслительной деятельности и потребностей общения, что обусловлено социальным характером практической деятельности" [146, с.40]. Рассматривая все разновидности указанного двуединства, она строит иерархическую систему, состоящую из трех ярусов и двух иерархических подсистем, закрепленных непосредственно за коммуникативно-мыслительной функцией и номинативной функцией языка – коммуникативной и номинативной подсистемами [146, с.40-44].

Такие виды мыслительной и коммуникативной функций, как двуединство – функция познания, национально-культурная функция (или этническая) и функция владения общественно-историческим опытом (или аккумулятивная), Л. А. Киселева относит к высшему ярусу (ярус I/в). Частные функции двуединства, такие, как, интеллектуально-информативная и прагматическая, в ее системе занимают ярус II. Далее, разновидности яруса II: эмоциональная, эмоционально-оценочная, эстетическая, побудительная, экспрессивная и контактная (или фатическая) представляют еще более низкий ярус – ярус III.

Ориентируясь на формализацию коммуникации "человек–компьютер", рассмотрим подробнее основную функцию языка – коммуникативную.

В упомянутой схеме Л. А. Киселевой предложения (простые и сложные) отнесены к I низшему ярусу, а сверхфразовые единства (СФЕ), а также единицы, находящиеся в иерархических отношениях с СФЕ, – сложное синтаксическое целое (ССЦ) и абзац формируют ярус II подсистемы. Единство абзацев, представленное в параграфе, разделе, главе, части, принадлежат к III ярусу. И наконец, к IV – высшему ярусу относится "текст" – коммуникативная единица высшего ранга. Она отмечала: "Приведенные коммуникативные единицы всех ярусов как единицы коммуникативной подсистемы языка имеют свои модели, схемы, типизированные способы построения и типизированные, узуальные средства, заполняющие эти модели" [146, с.141]. По ее мнению, эта подсистема реализуется не во всех сферах общественного употребления языка и не во всех формах его существования в полном наборе ее единиц [146, с.145].

Говоря о единицах языка подробнее, следует заметить, что это тот материал, который образует вторичную материальную систему, т. е. "в которых материальные элементы значимы для системы не столько в силу своих субстанциональных свойств, сколько в силу приписанных им свойств..." [261, с.19]. В процессе

функционирования языка в результате комбинации единиц языка образуются единицы речи. Из такого утверждения, т. е. речь есть результат функционирования языка, неизбежно следует, что в речи обнаруживается именно то, из чего состоит язык как средство общения – единицы языка всех видов и правил их комбинаторики [261, с.145 и т. д.].

Как известно, такие составляющие, как фонема, морфема, слово, относятся к единицам языка. Статус этих трех величин вызывает среди ученых много научных споров. Споры идут по поводу того, являются ли они единицами языка и какая из них признается основной языковой единицей. Разрешение этих вопросов, на наш взгляд, зависит от четкости требований, которым должны отвечать единицы языка. В основе таких требований, конечно, должна лежать идея, заложенная Ф. де Соссюром о том, что языковая единица, или языковой знак по Соссюру, – это "отрезок звучания, который, будучи взятым отдельно, т.е. без всего того, что за ним следует в потоке речи, является означающим некоторого понятия" и соответственно представляет собой "двустороннюю психическую сущность" [260, с.99], в которой означаемое (понятие) и означающее (акустический образ) связаны произвольной ассоциативной связью. Действительно, как пишет Л. Ельмслев, "язык организован так, что с помощью горстки фигур и благодаря их все новым и новым расположениям может быть построен легион знаков" [98, с.305].

Единицы языка должны отражать "существенные общие признаки языка" и тем самым иметь смысловые или эмоциональные значения и быть воспроизводимыми в готовом виде [262, с.13]. Многие ученые-лингвисты, такие, как В. В. Виноградов, А. И. Смирницкий, В. М. Солнцев и др., поддерживая точку зрения Ф. де Соссюра, считают, что слово есть "нечто центральное во всем механизме языка" [259, с.111], признают за "словом" статус основной языковой единицы [63, с.15, 261, с.184; 262, с.13-21].

Не вдаваясь в подробности, заметим, что в работах дескриптивистов и ученых других лингвистических школ указанные языковые единицы – фонема, морфема, слово получили иную оценку, отличающуюся от отмеченных.

Из требований воспроизводимости языковых единиц вытекает еще одно важное свойство единиц языка – их вычленимость, которое является следствием дискретности и неоднородности единиц. Следование принципу вычленимости, по мнению А. С. Мельничука, помогает не смешивать понятие единиц языка со многими другими лингвистическими понятиями [194, с.162].

Таким образом, в процессе функционирования языка его единицы должны: а) выражать некоторый смысл либо участвовать в его выражении и дифференциации; б) быть выделимы или вычленимы в качестве некоторых объектов; в) быть воспроизводимы в готовом виде и др. [261, с.185-186].

Переходя к изложению понятия "речь", отметим, что "исторически факт речи всегда предшествует языку" [260, с.57]. Утверждая первичность речи как явления непосредственно наблюдаемого, Ф. де Соссюр тем самым подчеркивал вторичность языка как выделяемого из речи, т.е. речь есть форма существования языка, а язык – абстрактная сущность.

Следует заметить, что в современной лингвистике под речью понимается не только устная и письменная речь, но и так называемая "внутренняя речь", т.е. мышление с помощью языковых средств (слов и т.д.), осуществляемое "про себя", без произнесения вслух [189, с.8]. Речь есть особая сфера, которая существует благодаря функционированию языка. Вследствие этого в речи можно обнаружить

все то, из чего состоит язык как средство общения, – единицы языка всех видов, а также всевозможные правила их комбинаторики. Об единицах речи можно говорить, как о специфических явлениях, главная особенность которых состоит в том, что они являются результатом свободного комбинирования единиц языка. Раскрывая сущность речи, Ф. де Соссюр писал: "Речь – сумма всего того, что говорят люди, она включает: а) индивидуальные единицы, зависящие от воли говорящих; б) акты фонации, равным образом зависящие от воли говорящих и необходимые для реализации этих комбинаций".

Следовательно, в речи нет ничего коллективного: "проявления ее индивидуальны и мгновенны" [260, с.57].

В лингвистике нет единого мнения о роли предложения. В лингвистических источниках существуют множество определений предложения, где одни ученые относят предложения к языку, а другие – к речи [120]. Мы будем придерживаться определения, предложенного В. А. Звегинцевым: "Предложение есть минимальная и вместе с тем основная единица речевого общения. ...Процесс речевого общения представляет обмен мыслями, но не обмен морфемами, словами и словосочетаниями" [120, с.172].

Основное отличие предложения от всех видов слов и словосочетаний – это то, что оно есть продукт, результат речевой деятельности, т.е. цельная речевая система, несущая информацию. Предложение отличается от указанных единиц и в грамматическом отношении, характеризуясь категориями модальности и предикативности. Согласно признаку производимости и сложности, предложение всегда можно противопоставить единицам языка [261, с.152].

Таким образом, само обозначение "язык – речь" подчеркивает единство (двуединство) языка и речи, предполагающее их органическую взаимосвязь и взаимодействие.

Ориентируясь в дальнейшем на обработку письменной речи, рассмотрим детальнее основные особенности этого вида речи.

1.2. Текст как основной объект современной лингвистики

Текст как объект текстологии и стилистики привлекал внимание ученых и раньше. Текстологи в основном занимались литературно-филологическим изучением древних текстов. А стилисты в целях изучения стилистических (экспрессивных) эффектов, достигаемых в результате искусного использования авторами лексических и грамматических элементов языка, в качестве объекта исследования рассматривали, как правило, художественные произведения.

К задачам теории литературы и литературоведения, как считает О.И.Москальская, относятся макрокомпозиция художественного произведения, развитие основных сюжетных линий произведения, их переплетение и расположение относительно друг друга [199, с.78].

Языковеды прошлого рассматривали объектом лингвистического изучения готовые тексты. При этом исследование текста было направлено, прежде всего, на анализ данного языкового материала. Однако, как правильно отмечает А. Е. Супрун: "Вопрос о том, как возникли эти тексты, оказывался за пределами интереса языковедов. Он мог интересовать текстологов или историков литературы, но с другой, нелингвистической точки зрения. ..." [263, с. 118-119].

В этой связи необходимо отметить и заслуги представителей системно-структурной парадигмы, которые, занимаясь описанием все более высоких уровней языковой системы, создали базу для изучения нового объекта – текста в его статике и динамике. Но эти исследования текста сводились в основном к внутриязыковому анализу, описание семантики языковых единиц ограничивалось пределами языка, без учета экстралингвистических факторов. Семантическая теория описывалась вне связи со знанием о мире его носителей, т.е., по существу, без учета человеческого фактора. По этой причине возникла необходимость создания адекватной теории значения, которая эксплицировала бы способность человека понимать данный язык и использовать его осмысленно в коммуникации.

Новая область языкознания – лингвистика текста, возникшая на основании идеи необходимости принципиального различения языка и речи, рассматривает текст уже под другим углом зрения. По Е. А. Реферовской: "Предмет ее изучения – построение, порождение текста, его конструктивные элементы с точки зрения как формальной, так и содержательной структуры формирования его смысла" [240, с.38]. Р. Барт считал, что предмет лингвистики текста – это "любой конечный отрезок речи, являющийся некоторым единством с точки зрения содержания, передаваемый в целях коммуникации и имеющий соответствующую этим целям внутреннюю организацию" [330].

По формулировке Р. Харвега, "структурный анализ текста строит перечень закономерностей на базе заданного материала, а лингвистика текста пытается найти текстообразующие закономерности, присущие всем текстам" [337, с.10]. Одной из таких ведущих закономерностей Е.А. Реферовская предлагает считать коммуникативное членение, так как без его учета невозможна сама лингвистика текста [240, с.27].

В целом, задача понимания текста естественного языка рассматривается обычно как получение семантического представления, а задачей генерации текста является по заданному семантическому представлению получение соответствующего текста на естественном языке.

Текст, являясь сложным многоплановым и многоуровневым образованием, в котором участвуют различные элементы и различные системы их отношений [182, с.115], многими исследователями не всегда понимается одинаково. И как всякий новый объект исследования в разных источниках он определяется по-разному. Для полной ясности проанализируем некоторые определения, которые, на наш взгляд, более объективно раскрывают суть понятия "текст".

И. Р. Гальперин в книге "Текст как объект лингвистического исследования" рассматривает "текст" как сложное и разностороннее явление. При определении системных, онтологических и функциональных свойств текста он считает необходимым учитывать такие его стороны, как размытость, неопределенность, стохатичность и детерминированность [70, с.4]. Он предлагает следующее определение: "Текст – это произведение речетворческого процесса, обладающее завершенностью, объективированное в виде письменного документа, литературно обработанное в соответствии с типом документа; произведение, состоящее из названия (заголовка) и ряда особых единиц (СФЕ), объединенных разными типами лексической, грамматической, логической, стилистической связи, имеющее определенную целенаправленность и прагматическую установку" [70, с.18]. С другой стороны, И. Р. Гальперин считает текст графическим отображением "кусочка действительности" или как порождение письменного варианта языка [70, с.5].

М. Пфюце под текстом понимает "определенную организованную по цели и смыслу совокупность фраз или фразовых элементов, между которыми имеются значимые отношения и функции, т. е. структурированное единство, представляющее в сознании в виде лингвистической единицы какое-либо комплексное явление действительности в его относительно законченной смысловой целостности" [347]. Очень удачно, на наш взгляд, определяет понятие "текст" ван Дейк, высказывая мысль о том, что "текст – это некая макроструктура, некое глобальное единство, не являющееся результатом сложения составляющих его микроструктур, а опирающееся на единство содержания" [334]. По А. А. Леонтьеву "текст есть функционально завершенное речевое целое" [176, с.28]. С точки зрения языкового знака "Текст – это первичный переменный языковой знак отрезка речи (речемы), выражающего одно или множество умозаключений (логических единств, рядов суждений)" [132, с.4] или под текстом понимается любая семиотически организованная последовательность знаков (Б. А. Успенский); речь или текст есть система, несущая информацию (В. М. Солнцев) [251, с.18]; текст представляет собой линейно упорядоченную совокупность дискретных знаковых единиц разного объема и сложности [309, с.142]. Рассматривая текст как комплекс языковых средств и мыслительного содержания, можно "тексту" дать и такое определение: "Текст – это не только последовательность материальных знаков, но и все то, что стоит за ней и образует своего рода глубинную перспективу. Поэтому под текстом надо понимать комплекс как языковых средств, так и мыслительного содержания, возникающий в процессе понимания" [182, с.114].

Можно было бы продолжить ряд таких определений "текста", но анализ разнообразных их вариантов дает нам основание классифицировать их со следующих точек зрения: лингвистической [17, с.470; 70, с.18; 71, с.63-64; 120, с.170; 337, с.148; 341, с.385; 343, с.232-296], семиотической [144, с.121; 335, с.118; 350, с.154; 351, с.13], логико-лингвистической [133, с.48; 207, с.172 и др.], информационной [27, с.443-449; 320, с.98], структурной, прикладной и математической лингвистики [227, с.170]. Определения понятия "текст" приведены и во многих других работах, но все они, в основном, рассматривают текст как некое структурированное по определенным законам единство, состоящее из языковых единиц – предложений, объединенных между собой и образующих более крупные единицы – сверхфразовые единства (СФЕ), абзацы, параграфы, главы, разделы и т. д., служащие для передачи определенного законченного содержания, отвечающего целям коммуникации [69].

Известно, что если рассматривать текст на уровне предложений, то он состоит из смены одиночных и объединенных в группы несамостоятельных предложений. При этом, такое свойство текста, как линейность, "позволяет показать возможную роль предложения – минимальной коммуникативной единицы – в монологическом высказывании" [116, с.22].

К. Кожевникова, отнеся "текст" к целостной, замкнутой высшей коммуникативной единице, совершенно справедливо настаивает на том, что основным конституирующим признаком текста является связность семантическая, которая наличествует обязательно, и формальная, которая может эксплицировать семантическую применением грамматических и лексических средств языка [149, с.49-68].

Связный текст, передавая реальную внеязыковую связь действий, событий, ситуаций, оценок, подлежащих сообщению, естественно, должен определенным образом отражать и выражать эту связь. В связном тексте можно обнаружить

некоторые особенности его лексической организации. Развитие одной конкретной "темы" в какой-то мере обусловливает понятие слов определенной тематической группы на протяжении всего текста.

Помимо разнообразных дейктических элементов и различных средств собственно синтаксической связи (союзы, союзные слова и т.д.), связный текст характеризуется и такой грамматической организацией, как согласование форм времени, основанное на создании общей временной перспективы повествования или сообщения, и, как соотнесение субъектов действий и состояний, основанное на их различении и отождествлении и т. п. [321, с.21].

Следующей характеристикой текста является "неоднозначность связей единиц текста", которая связана с внутренними законами любого информационного процесса [229, с.17-21]. Неоднозначность связей между единицами текста можно понять, если мы обратимся к процессу создания текста авторами произведений. Автор текста, будучи источником сообщения (текста) и обладая определенным объемом знаний о внешнем мире, в силу своей лингвистической компетенции, создает некоторое описание реальной действительности. При этом, он руководствуется жизненным опытом и знанием языка, а также мотивационной направленностью, ориентацией на тип текста (рассказ, статья, учебник, переписка и т.д.), эмоциональным состоянием и т.д., и т.п. [150, с.152; 208, с.52, 53; 227, с.170-171]. С точки зрения семантики текста, в общем случае, все тексты делятся на два класса: тексты традиционной семантики и тексты, порождающие собственную семантику [322, с.154-155]. К первому типу относятся тексты описательного характера, т.е. большинство описательных научно-технических текстов, вся деловая проза [100], корреспонденция и т.д. Считается, что в таких текстах представлен референт, заранее существовавший (хотя бы в сознании автора). Все тексты, которые сами порождают свой референт, относятся ко второму типу текстов, порождающих собственную семантику. Сюда можно отнести тексты, описывающие математические конструкции, кодифицированные тексты юридических законов, тексты проектов, все художественные тексты [322, с.153-154; 253, с.306-482]. Относительно художественного текста Г. В. Степанов отмечает, что здесь автор посредством языка создает совершенно особую ситуацию, хотя и соотносимую с реальной, но не равную ей [264, с.73]. Из характера описательных текстов, т.е. текстов традиционной семантики, можно заметить, что содержание, вкладываемое в текст автором, совпадает с объективным содержанием, передаваемым семантическими структурами предложений и абзацев [66, с.50; 315, с.4].

С точки зрения лингвистики текста необходимо различение языкового и смыслового, понятийного и обобщенного содержания текста [240, с.156]. В этом отношении А. В. Бондарко пишет: "... план содержания текста структурирован лингвистически ..., а смысл текста структурирован понятийно" [52, с.100-113]. Таким образом, конкретное содержание какого-либо текста членится на составляющие его элементы, на части, главы, на отдельные отрезки текста и на СФЕ. В свою очередь, основное содержание СФЕ состоит из составляющих его ведущих предложений. При этом смысл текста не является совокупностью смысловых образующих его предложений. Каждое предложение имеет "невидимый" (т.е. лингвистически не выражаемый) подтекст. Выявление этого подтекста служит задачей пресуппозиций [234, с.29].

Смысл – это интеграция (синтез) конкретных значений, или, иначе говоря, интегративная (смысловая) связь двух речений. Грамматическая связь работает

внутри предложения в виде управления, согласования и примыкания словоформ. Смысловая связь работает между предложениями, связывая их по смыслу [105, с.87].

Смысл всего текста (всего повествования) представляет собой обобщение, заключение, вывод из его содержания. Общая идея, лежащая в основе всего текста, и представляет собой общий смысл текста.

Е. А. Реферовская в книге "Коммуникативная структура текста" пишет: "Смысловая структура текста находит свое отражение в его лингвистической структуре, получающей соответствующие лексико-грамматические характеристики, которые составляют предмет изучения лингвистики текста" [240, с.3]. Итак, исследование лингвистической структуры текста необходимо начинать с выявления основных единиц, из которых строится текст. При этом основное внимание необходимо обращать на языковые и коммуникативные структуры предложения, а также и других компонент, превышающих предложение и, наконец, текста в целом. Коммуникативное членение текста – психо-лингвистическая категория, отражающее как объективные и реальные связи и явления реального мира, так и субъективное отношение человека к реальным явлениям и связям действительности [240, с.9].

1.3. Текст и его единицы как знаки

Подходы к определению понятия "знак". Современная наука проявляет все более возрастающий интерес к изучению знаковых систем. Об этом свидетельствуют возникшие в последнее время целые разделы в философии, логике, кибернетике, лингвистике, психологии и других науках, в которых основным предметом исследования служат знаки и знаковые системы.

Изучением наиболее общих свойств знаков и знаковых систем занимается семиотика. Рассматривая наиболее общие закономерности знаковых систем, семиотика становится междисциплинарной наукой, оказывая действенное влияние на ход развития многих гуманитарных и естественных наук.

При семиотическом подходе к знаковым системам особо подчеркивается первичность естественного языка как основы коммуникации и сотрудничества для более широкого класса языков, кодифицирующих обозначаемое, а также зависимость искусственных знаковых систем от естественного языка.

Современный мир характерен тем, что по мере развития общества, усложнения форм общественной жизни, обогащения и развития сознания развиваются и усложняются формы и виды общения. Следовательно, развиваются и усложняются естественный и искусственные языки [251, с.15]. Этот факт говорит о том, что знак и знаковая деятельность становятся существенным элементом социального прогресса и жизнедеятельности человека. Такие процессы, как управление, накопление и переработка информации на основе широкого использования электронно-вычислительной техники, требуют решения ряда задач, связанных с сознательным конструированием новых знаков и знаковых систем. В этой роли естественный язык, несмотря на свою универсальность и гибкость в процессе коммуникационной деятельности людей, пока значительно уступает искусственным языкам, которые являются наиболее удобными средствами моделирования и замещения естественного языка и создаются на основе такого языка. Здесь уместно

привести лаконичную формулировку Е. Н. Панова: "Естественный язык человека осуществляет семиотическое моделирование и тем самым служит универсальной семиотической матрицей, на которой можно вторично построить практически неограниченное число самых разнообразных знаковых и информационных систем" [224, с.87]. В доказательство можно назвать ряд действующих автоматизированных информационных систем (АИС) и информационно-поисковых систем (ИПС), построенных на основе законов знаковых систем. К тому же программная обеспеченность современных компьютеров позволяет реализовать различные информационные системы любой сложности.

В последние годы в языкознании одним из основных объектов исследования становится текст. Такое направление в выборе объекта исследования обусловлено как социальными потребностями, так и логикой внутреннего развития самой лингвистической науки. Отмечая подобное явление и в области применения ЭВМ в разных сферах общественной жизни и особенно в информатике (т.е. в смысле науки, изучающей процессы передачи и переработки информации), А. П. Ершов писал: "Если расчетные инженерно-научные и экономические применения ЭВМ предопределили облик вычислительного дела в 50-е и 60-е годы, то сейчас центр научной проблематики смещается в сторону обработки текстовой информации в широком смысле, а также массового применения ЭВМ" [100, с.5]. Такое смещение научной проблематики тесно связано, во-первых, с задачами использования ЭВМ для свертки, хранения и воспроизведения информации, или с задачами поиска публикации по каким-то признакам этой информации. Во-вторых, это смещение определяется задачами реферирования, аннотирования научных статей, перевода иностранной литературы и, наконец, связано с новыми проблемами компьютерного обучения в школах и вузах. Все процессы, реализующие указанные задачи, имеют дело с вводом, обработкой и выводом текстовой информации.

Тем самым язык приобретает новые функции. Однако, задачи, связанные с конструированием знаков и знаковых систем, требуют детального учета основных понятий семиотики, глубокого понимания принципов функционирования языка вообще и казахского в частности.

В отечественной и зарубежной литературе изложены различные теории, раскрывающие суть знака и учитывающие разнообразные их типы и виды [2, 81, 177, 195, 229, 260, 289]. Не вдаваясь в подробности философских суждений о природе знака вообще и языкового знака в частности, отметим, что все приводимые определения философов в основном сводятся к таковым, приведенным "Философском словаре": "Знак – материальный предмет (явление, событие), выступающий в качестве представителя некоторого другого предмета, свойства или отношения и используемый для приобретения, хранения, переработки и передачи сообщений (информаций, знаний)" (ФЭС, 1983, с.191).

Двойственная природа знака позволяет человеку отличить знак от "не-знака". Такая природа знака дает нам возможность различать материальную форму и значение, в силу которого он становится заместителем обозначенного предмета.

Проблема значения знака интересует многих представителей различных областей знания. В отличие от материальных свойств знака, "значение" его является функциональным свойством знака особого рода, т.е. идеальным функциональным свойством. Такое понимание природы значения знака позволяет раскрыть одну из основных знаковых антиномий, а именно находится ли значение в знаке или вне его. Принципиальные расхождения во взглядах как зарубежной, так и отечественной

семиотической науки на понимание сущности знака и значения в основном можно объединить в два направления – ментализм и антиментализм. Согласно первому, значение, семантика, идеальное и т.д. включаются в структуру знака и языка. В соответствии со вторым направлением, в звуковом комплексе (и соответственно в видеографической системе) никакого идеального содержания нет, т. е. все, что подразумевается под значением, семантикой и т.д., находится за пределами знака (в нейродинамической системе субъекта). Последний подход к природе знака и значения считается наиболее адекватным, так как раскрывает действительный смысл знака как субстита (заместителя тех мыслительных процессов, которые происходят в мозгу субъекта) [298, с.123-124].

Обобщая понятия о сущности значения, следует отметить, что значение знака возникает и существует в процессе общественной познавательной деятельности, которая осуществляется внутри сложной связи человека с другими объектами материального мира, со средствами отражения объективной реальности и т.д.

Следующее свойство знака связано с тем, что форма знака в знаковых системах конвенциальна, т.е. она является условной, произвольной по отношению к обозначаемому предмету. Это утверждение дает основание полагать, что знак не может заменить материального тела обозначаемого, а способен заместить лишь некоторые его функции. Конвенциальность или произвольность знака в знаковой системе Ф. де Соссюр считал одним из главных принципов, на который необходимо опираться в изучении языковой системы, т. е. такой системы знаков, которая не обусловлена свойствами объекта. Он писал: "Основной принцип произвольности знака не препятствует различать в каждом языке то, что в корне произвольно, т. е. немотивировано, от того, что произвольно лишь относительно. Только часть знаков является абсолютно произвольной; ... знак может быть относительно мотивированным". По мнению Ф. де Соссюра, изучение системы знаков с точки зрения ограничения произвольности языкового знака – "наилучшая основа исследования" [260, с.163-165].

Рассмотрение знаковой системы на основе функционального подхода дает нам возможность выявить природу знака, а также проследить в онтогенезе становление знаковой функции. Это позволяет использовать в качестве знака различные предметы и явления – заместители реальных явлений, формирование способности употреблять одни объекты для представления других [251, с.12]. Иными словами, материальный предмет становится знаком в результате того, что он превращается в элемент некоторой системы отношений, в рамках которой он играет особую роль, выполняет специфическую – репрезентативную роль, а не в силу природной определенности.

Рассматривая язык как систему знаков, В. В. Налимов подразделяет ее на жесткую и мягкую. При этом, он относит знаковую систему к очень жесткой, когда знак и обозначаемое связаны каким-то безусловным образом. Систему знаков, как считает В. В. Налимов, можно отнести к мягкой структуре (достаточно сложной), когда правила приписывания смыслового содержания знакам или их комбинациям не подчиняются четкому упорядочиванию. При такой дифференциации обыденный язык будет относиться к мягкой системе, в силу отсутствия однозначного соответствия между знаком и обозначаемым. Примером может служить тот факт, что при некоторых обстоятельствах один знак употребляется для обозначения того, что обычно обозначается двумя антисинонимичными знаками [201, с.67-69].

Таким образом, естественный язык относится к вторичной семиотической системе и мягкой структуре знаковых систем.

Из различных семиотических трактовок языкового знака (логическая, прагматическая, лингвистическая) особый интерес для нас представляет инженерно-лингвистический подход к знаковой ситуации Р. Г. Пиотровского с соавт. [177, с.22-43]. При разработке теории знака Р. Г. Пиотровский последовательно опирался на известные подходы к определению сути знака рядам ученых [155, 259, 260, 265, 316 и др.]. Согласно данной теории, в означаемом референте содержится, по крайней мере, два самостоятельных компонента – денотат и десигнат. Первый из них, присоединяясь к имени, является целостным отражением референта, где имя отражает некоторый единичный объект внешнего мира. Второй компонент – десигнат (интенсионал, концепт, сигнификат, смысл) представляет собой смысловую единицу, выражающую концепт, общую сущность всех предметов, принадлежащих к классу, обозначаемому данным конкретным именем. Развивая эту теорию, Р. Г. Пиотровский вводят в структуру знака третью компоненту – коннотат, который над денотатом и десигнатом каждого знака надстраивает вторичные смысловые оттенки и оценки – эмоционально-оценочные аспекты означаемого [177, с.33-34]. Таким образом, если у Г. Фреге, который вошел в историю науки как зачинатель семиотических исследований, "знаковая ситуация" схематично изображается в форме треугольника, где его вершины означают знак, денотат, смысл, то у Р. Г. Пиотровского (и его соавторов) соотношение между именем и референтом может быть представлено как на *рис. 1.*

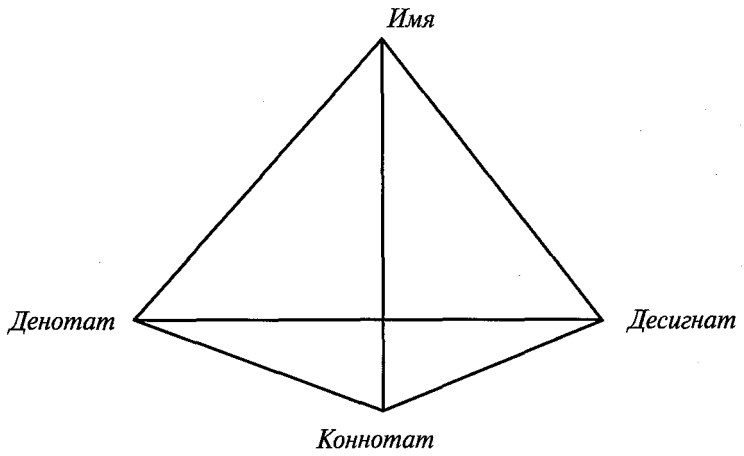


Рис. 1. Схема слова-знака

В своих последующих разработках теории знака Р.Г. Пиотровский рассматривает знак как элемент динамической системы, т.е. с точки зрения функционального подхода к определению знака [344, 353, с.51-58]. Такой подход к лингвистическому знаку позволяет обнаружить план выражения, план содержания и план интерпретации сообщения потребителем [229, с.17].

Детализируя эти отношения, Р.Г. Пиотровский включает в знаковую ситуацию (семиозис) шесть компонентов, определяя при этом знак как многокомпонентную психическую сущность: 1) референт (предмет); 2) сигнальный референт; 3) денотат (образ предмета); 4) десигнат (концепт, смысл, понятие); 5) коннотат (эмоционально-оценочный аспект означаемого); 6) имя (означающее).

Здесь под референтом следует понимать материальные предметы, их классы, созданные нашей фантазией отвлеченные понятия, а также их характеристики связи [177, с.23].

Ориентируясь на использование языковых знаков в динамике (речевой деятельности), рассмотрим, как происходит последовательный переход от семиотической схемы Г. Фрега к динамической модели знака Р. Г. Пиотровского для различных типов языковых знаков.

1.4. Предложение как речевой знак

Огромный интерес ученых к изучению семантики предложения отмечается с середины 60-х годов. При этом проблема рассматривалась и с позиции лексикологии, и по линии выявления ситуативных значений, и со стороны логики, и, конечно, с позиции лингвистики [13, с.7]. С какой-бы позиции не рассматривалась семантика предложения, основной целью исследователей являлось определение отношений между высказыванием и обозначаемой им экстралингвистической ситуацией или событием.

До последнего времени особое внимание языковедов было обращено на описание предложения как такового, т.е. вне всякого контекста. Такой подход отмечался и при привлечении методов, синтаксической парадигматики, синтаксической деривации, и трансформационных методов. Основной причиной такого явления была недостаточная изученность плана содержания в отличие от плана выражения.

В настоящее время акценты в изучении семантики предложений сместились в сторону контекстного описания предложения с помощью правил контекстно-связного синтаксиса. Этот факт свидетельствует о том, что первостепенную роль в изучении предложения начинает играть план содержания [44, с.25].

В результате многих исследований в области лингвистики текста выяснилась несамостоятельность предложения, т.е. его зависимость от единиц более высокого уровня, таких, как абзац, сверхфразовое единство, или от всего текста. Предлагается даже в пределах текста говорить не о предложении, а только о псевдопредложении [120, с.185-188]. Это подтверждает факт контекстуальной зависимости предложения, т.е. зависимость предложения от множества факторов, лежащих за пределами самого предложения [44, с.26].

В целях установления статуса предложения некоторые исследователи предлагают принять предложение как максимальную единицу языка и минимальную единицу речи [240, с.36], а также как особую единицу языка, так как конкретное содержание каждого из них своеобразно и индивидуально, что требует особых подходов и приемов к изучению его семантики [232, с.86]. А. Е. Супрун предлагает рассматривать предложение как единицу грамматической организации, как грамматическую модель высказывания [263, с.52]. Соотносясь с высказыванием как речевой единицей, предложение, так же, как и слово, выступает номинативной единицей языка. Причем предложение, будучи языковым знаком, несет номинативную функцию в одно - или многословной форме. Соотнесение предложения с высказыванием и суждением дает основание подвергнуть его анализу на трех уровнях: логико-грамматическом, логико-коммуникативном и синтаксическом (традиционном).

В лингвистике имеются различные определения термина "предложение". Приведем только некоторые из них. В книге "Основы построения описательной грамматики современного русского языка" дано следующее определение: "Предложение есть синтаксическая структура коммуникативного назначения, в основе которой лежит одна из существующих в системе языка конструктивных форм, обладающих категорией предикативности как парадигматически выявляющегося сложного единства категорий синтаксического времени и объективной модальности" [222, с.137]. В. А. Звегинцев считает: "Предложение – это словосочетание плюс какой-то добавок, некая почти магическая формула, которая наделена властью преобразования". Здесь "какой-то добавок" – это не что иное, как значение предложения [120, с.134]. Уместно отметить, что, как и любые значимые единицы языка, значение предложения определяется, во-первых, теми отношениями вне языковой действительности, выражением которых и является язык и его синтаксическая структура. Во-вторых, теми отношениями, в которых данное преложение связано общностью значения и совпадением какой-то части составляющих его компонентов [321, с.102]. Единицы языка разных уровней, в том числе и структурная схема предложения, обладают семантическими особенностями, создаваемыми взаимодействием в них категориального и конкретного значений [317, с.166-167]. В. А. Звегинцев отмечает, что "если язык есть система средств для передачи в процессе общения некоторого содержания, то эта система организуется вокруг значения" [120, с.115].

Выяснение роли предложения в системе языка и речи, его семантическая и синтаксическая организация все еще остаются проблемными задачами языкознания, логики и психолингвистики. Многие исследования ученых в этой области раскрывают различные подходы к решению всех сложных проблем, касающихся роли и места предложения в языковой системе [5, с.112-204; 246, с.83-87]. Мы не будем подробно останавливаться на этой проблеме, а только отметим некоторые основные моменты.

За основу предложения чаще всего принимают [328, с.48-49]: 1) внеязыковую ситуацию (референтная концепция); 2) структуру мысли в виде предметно-атрибутного суждения или в виде пропозициональной функции (логическая концепция); 3) verbum finitum и его грамматические категории (морфологическая концепция).

Рассмотрим несколько подробнее референтную концепцию семантической основы предложения. В соответствии со знаковой природой единиц языка, под смыслом предложения понимается отражение и обобщение в сознании человека некоторого фрагмента действительности, представленного в виде некоторой простейшей ситуации или события [73; 88, с.52; 137, с.141, 142; 151, с.52; 211, с.297; 267, с.15; 311, с.15]. Такую простейшую ситуацию в литературе называют "психической ситуацией" [178], а в "смысле" при этом выделяют денотативную и десигнативную составляющие.

Считая высказывание полным языковым знаком, В. Г. Гак полагает, "что референтом высказывания является ситуация, т.е. совокупность элементов, присутствующих в сознании говорящего в объективной действительности, в момент "сказывания" и обусловливающих, в определенной мере, отбор языковых элементов при формировании самого высказывания" [74, с.358].

По определению В. С. Храковского, "смысловая структура предложения представляет собой вырезанный и обработанный мыслью и языком фрагмент

действительности, который принято называть индивидуальной денотативной ситуацией или событием" [301, с.5]. Во многих исследованиях смысловая структура предложения моделируется с опорой на грамматические понятия и категории, т.е. "поиски соотнесенности в предложении грамматических признаков с семантическими" [121, с.17]. Считая, что "предложение как акт, в котором реализуется формирование, выражение и сообщение мысли, не сводимо к развертыванию цепочки слов", Г. А. Золотова представляет "структурно-смысловыми элементами предложения" понятия производителя действия, предмета исчисления, самого действия, состояния, качества, времени, причины и т.д. [122, с.90].

Денотатом предложения считается некоторая конкретная ситуация, отраженная в нашем сознании, а десигнат – некоторое представление об определенном классе ситуаций, которое формируется в сознании человека путем отвлечения от множества однотипных ситуаций [75, с.232; 267, с.9]. Но, воспроизводя конкретное предложение, автор может выражать в нем не только конкретное событие, а также и нечто общее, что всречается в однотипных событиях действительности. В содержательном аспекте предложения, кроме уже отмеченных, выражаются также разные эмоциональные, экспрессивные и оценочные оттенки. Как считает Л. М. Васильев, содержание предложения как высказывания есть его речевой смысл, который включает в себя три аспекта: номинативный, модальный и коммуникативный. Здесь модальный план предложения рссматривается как конкретная реализация семантических категорий модальности, темпоральности. Модальные и номинативные планы предложения (высказывания), вместе взятые, составляют его семантическую структуру, его денотативную и сигнификативную семантику, а коммуникативный план имеет дейктическую природу [58, с. 5-10].

Изучение плана содержания языковых единиц требует учитывать одновременно все три основополагающих фактора, которые участвуют в формировании семантики: форму языковых единиц, денотат и сигнификат [232, с.89].

В рамках высказываний семантические элементы могут быть подразделены на три типа по признаку отражения ими того или иного ареала объективной действительности. Первый тип – семантико-синтаксические, т.е. которые отражают ситуации объективного мира, находящегося за пределами человека; второй тип – интенционально-семантические, т.е. те, которые отражают речевую и психическую деятельность человека; третий тип – логико-семантические, т.е. те, которые отражают мыслительную деятельность человека [302, с.11].

Необходимо особо отметить, что в структуре предложения-знака обязательным компонентом является прагматический аспект. Этот факт можно понимать шире коннотативного аспекта слова-знака, так как он, кроме коннотативной, включает еще и информативную, или коммуникативную составляющую. Информативность здесь понимается не в смысле истинности предложения и соответствия мира истинным условиям предложения, а в русле проблем актуального членения.

Таким образом, учитывая определенные синтаксические отношения между словами, входящими в предложение, мы представляем семантическую структуру знака-предложения в речи в виде *рис. 2.* Как видно, сложность знака-предложения выросла соответственно новым отношениям по сравнению со знаком-словом, т.е. с появлением информационного аспекта предложения *Рп* и соответствующими связями с другими компонентами знака.

Согласно представленной семиотической структуре-предложения, под смыслом предложения понимается отражение в сознании человека отношения линейной

последовательности имен предложения к отраженному в сознании событию (как совокупности, отмеченных выше компонентов означаемого).

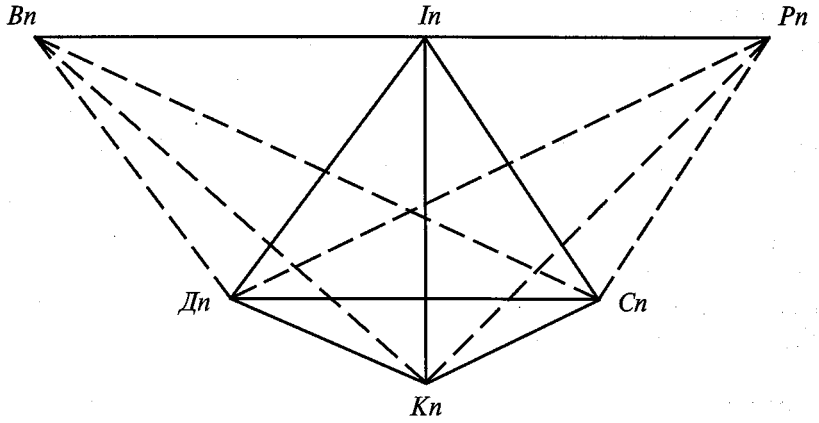


Рис. 2. Семиотическая структура знака-предложения. $D_п$ – денотат предложения; *Сп* – десигнат; *Кп* – коннотат; *Вп* – синтаксический аспект предложения; *Рп* – информационный аспект; *Іп* – цепочка из "*п*" имен предложения.

1.5. Сверхфразовое единство и абзац как знаки

Распространение объекта исследования на совокупность составляющих его элементов представляет собой основу любого научного анализа. Изучив первоначально структуру более простых языковых сущностей, например фонем и грамматических категорий, ученые стали применять этот научный прием на материале более значимых единиц языка [232, с.62-66].

Как известно, в структуре связного текста можно выделить его отрезки, имеющие разную длину. Эти отрезки находятся между собой в некоторой определенной связи. Разграничение их осуществляется автором или по содержанию, или формально. Формальное членение связного текста проявляется в виде абзаца, раздела, главы, параграфа и т.д. Имеется ряд работ, где выделяются различные признаки текстовых единиц, объединяющие предложение в более крупные текстовые единицы [152, 189 и др.].

В лингвистической литературе происходит много споров среди ученых о сущности сверхфразовых единств (СФЕ) и абзаца [192, 199, 240, 254, 303 и др.]. В указанных и в ряде других работ авторы относят абзац как единицу текста и к композиционной, и к логической, и к семантической единице [131; 171, с.46; 193, с.23; 205, с.206; 225, с.9; 268, с.208; 255, с.208; 269, с.156; 294, с.220; 306, с.155]. В этой связи рассмотрим некоторые суждения ученых, касающиеся природы СФЕ и абзаца. Следует заметить, что как синонимы СФЕ в различных научных источниках встречаются и такие термины: сложное синтаксическое целое (ССЦ), структурно-смысловое единство и логико-синтаксическое единство.

Т. И. Сильман по поводу природы абзаца пишет: "Сочетание предложений, которыми заканчивается один абзац и начинается другой, не имеет достаточно отчетливо выраженных грамматических границ" и "от одного абзаца к другому ...

имеется некий разрыв, некий скачок мысли" [256, с.106-107]. Считая абзац некоторым отрезком текста и как некое единство содержания, он представляет абзац как синтаксическую структуру или как синтактико-интонационное единство, состоящее из одного или нескольких предложений, соединенных между собой синтаксическими связями, пронизанных лексико-местоименными повторами и объединенных единой общей темой [256, с.112].

Н.А. Турмачева замечала, что "... сложное синтаксическое целое (ССЦ), являясь смысловым и грамматическим единством, есть категория семантико-синтаксическая..." и "... абзац, выделяемый в соответствии с композиционными и экспрессивно-стилистическими задачами, является категорией стилистической, точнее стилистико-семантической" [287, с.177]. Стремление ученых выяснить природу СФЕ и абзаца приводит и к выявлению их общности в плане семантического содержания. Например, Е. А. Реферовская считает, что общими между СФЕ и абзацем являются то, что оба могут содержать законченный тематический отрезок текста, служить продвижению основной линии повествования, включать побочные вставки и т. д. К отличительным характеристикам она относит то, что СФЕ содержит высказывание, образующее объективно ограниченную микротему, когда как для абзаца такое явление не обязательно. Она не исключает, что в ряде случаев абзац может содержать целостную микротему, тем самым формально совпадать со СФЕ [240, с.50-54]. Основной особенностью СФЕ и абзаца, по ее мнению служит, то, что членение текста на абзацы носит субъективный характер, т. е., относясь к композиционно-стилистическим приемам членения текста, абзац связан с личным вкусом и индивидуальной манерой автора подчеркивать основные моменты его повествования. СФЕ, в отличие от абзаца, представляет собой объективный способ понятийно-лингвистического членения текста [240, с.55].

Логико-синтаксическое единство, В. Д. Ившин относит к номинативной единице языка и считает, что "оно структурно объединяет одной общей мыслью два и более предложений, в речи соотносится с абзацем, выражающим ряды суждений" [132, с.3-8].

Вслед за С. И. Гиндиным под СФЕ И. Б. Хлебников и Н. Ф. Потапова понимают "структурно-смысловое единство, где ведущей является семантическая связность" [299, с.44], а "синтаксическая выступает как средство поддержания семантической" [80, с.134].

Многие исследователи подчеркивают "стилистическую значимость" абзаца, рассматривают абзац как "производное и детерминат" индивидуального стиля и т.д., и т.п. (Т. И. Сильман, В. А. Кухаренко, Т. С. Ряднова, В. П. Николаева, И. М. Гриценко, Н. Е. Дорофеева и др.) [270, с.108]. В этой связи Л. Г. Каминскене подчеркивает, что членение художественного текста на абзацы представляет собой часть художественного замысла писателя и поэтому редактированию или усовершенствованию не подлежит [138, с.5]. Взяв за основу исследования художественные тексты, Н. И. Серкова утверждает, что, хотя в членении речи на абзацы в художественном тексте отсутствует какая-либо общепринятая схема, нельзя отрицать присутствие в прозе всякой внутреней меры членения. По ее мнению, "членение прозы ориентировано на конкретно-историческое представление о мере, балансе, гармонии, пропорциях и т.д., выработанных данным художественным направлением на основе соответствующих эстетических категорий" [270, с.115]. Существование абзаца в художественной прозе Н. И. Серкова объясняет результатом

единства объективного (художественная норма) и субъективного (содержание–намерение). Ею все же при квантовании художественного текста предпочтение отдается субъективной стороне – содержанию-намерению, она считает его главной "мерой" [270, с.120].

Многими учеными отмечаются "преимущества" и "недостатки" в определении границ СФЕ и абзаца в связном тексте и в установлении их статуса как единицы языка. Так, по отношению к СФЕ, абзац имеет бесспорное преимущество, обладая четкой и однозначной выделяемостью, а именно его расположением между двумя красными строками. В отношении СФЕ Л. М. Лосева, например, пишет: "Вообще дать формальные признаки конца сложного синтаксического целого довольно трудно, если не сказать больше, что просто невозможно" [183, с.60-68; 199, с.32-34; 240, с.69-81].

Исследователями структуры абзаца обычно указываются следующие "слабые места" абзаца: а) субъективный подход автором членения текста на абзацы; б) не все абзацы выражают завершенную мысль; в) абзацы не имеют стабильной структурной модели, что не позволяет считать их единицей языка; г) абзацы имеют колеблющийся объем (от одного предложения до нескольких страниц) [199, с.31-32; 179].

Чан Нгок Тхем в статье "Абзац как грамматико-семантическая сущность" делает попытку "защитить" статус абзаца как языковой единицы, опровергая перечисленные "слабые места" абзаца. Классифицируя абзацы на структурно завершенные (СЗ) и структурно не завершенные (СНЗ), а также на автосемантические СЗ и синсемантические СЗ, автор делает вывод, что "абзац как грамматико-семантическая сущность должен быть признан единицей языка в противопоставлении конкретному абзацу как проявлению речи" [303, с.118-123].

Одной из проблемных задач в исследовании внутренней структуры в СФЕ и абзацах является установление формальных и семантических связей между предложениями в указанных отрезках текста. По данному поводу ученые замечают, что формальные показатели связи предложений, образующих СФЕ, не всегда проявляются с достаточной четкостью, но часто такие показатели представлены в виде повтора одних и тех же элементов предложения. В этом же качестве используются союзы, вводные слова, наречия и т.д. [43, с.182]. При отсутствии формальных показателей семантическая связь между предложениями в СФЕ может обнаруживаться на основании логико-понятийных отношений, которые при всем их разнообразии могут быть сведены к двум основным типам: радиации и конкатенции [76, с.91]. Отношения радиации наблюдаются между теми предложениями, где изображаемые факты имеют связь по признаку пространства, отношения конкатенции – по признаку времени [232, с.92-93].

Известно, что были сделаны попытки развития навыков письменной речи на уровне самостоятельного сочинения (composition) в англо-американской дидактической текстологии. Авторы этих идей и экспериментов большое значение придавали СФЕ и абзацам, указывая, что в "правильно" написанном сочинении они должны иметь определенную последовательность и располагаться в тексте в соответствии с определенной моделью [332, с.11].

К этому проблемному вопросу организации большого отрезка текста (главы) с концептуальной точки зрения обращались и отечественные исследователи [352; 282, с.49-54].

Субъективный характер абзацев в различных связных текстах во многих случаях препятствует выдерживанию авторами произведений строго определенной

модели расположения абзацев в тексте. Осуществление последнего, конечно, привело бы к огромному облегчению усвоения прилагаемого письменного материала.

Таким образом, на основании изложенного обзора исследований можно сделать вывод о том, что из всех подходов, раскрывающих суть абзаца, наиболее распространен семантико-синтаксический. При этом основной единицей текста, объединяющей несколько предложений в тексте, является "сверхфразовое единство", или "сложное синтаксическое целое" [70, с.67-73; 183, с.61-84; 199, с.16-42; 226; 236]. Все определения этого понятия предъявляют в основном следующие требования: наличие смысловой и синтаксической (формальной) связи между предложениями в пределах данной единицы текста [189, с.8-11]. Но синтаксическая связь, для одних авторов "сложное синтаксическое целое", совпадает с абзацем [253-255, 271, 294], а для других оно может быть и меньше абзаца и состоять из нескольких абзацев [22, 82, 116, 202, 205, 272].

В качестве основного семантического компонента текста в нашем исследовании выбран абзац. Для обоснования такого выбора вначале следует отметить, что вся окружающая нас материя существует в виде разнообразных конкретных объектов и систем, которым свойственны движение, структурность, связь и пространственно-временное взаимодействие, а также наличие множества общих и частных признаков. Несмотря на обилие таких свойств, познание этой материи осуществляется путем квантования действительности и отражения этих квантов в виде определенных спаянных единством словесных блоков [126, с.34]. Основываясь на различных исследованиях лингвистов, психологов и логиков, считаем возможным принять для письменного текста в качестве такого словесного блока абзац.

Следует отметить, что говорящий и слушающий интуитивно выделяют в составе текста определенную иерархическую сеть тем, подтем и микротем [57, с.201; 106, с.15; 189, с.27; 307, с.107]. Обычно минимальное число объединенных простых сообщений, которое может нести одну или несколько мыслей, называют микротемой [126, с.35; 183, с.60]. В случае, когда микротема состоит из двух или более мыслей, считается необходимым существование связи в виде определенных отношений (временных, пространственных и др.) в соответствии с предметными свойствами их референтов. Итак, тема всего текста – это объединение нескольких микротем, связанных между собой определенными отношениями (временными, пространственными, функциональными и т.п.) в соответствии с предметными свойствами референтов, общими для всех микротем.

Проведенные различные эксперименты ученых по сегментации текста показали, что испытуемые для передачи мыслей группируют их в тематические группы предложений, которые, как правило, совпадали с абзацами в тексте [51, 269, 273, 342]. Из результатов таких экспериментов, полученных как у нас в стране, так и за рубежом, вытекает естественный вывод: выделение групп предложений в абзацы не является произвольной операцией автора текста, а основано стремлением его квантовать смысловое содержание. С другой стороны, ясно, что деление текста на абзацы способствует лучшему усвоению и запоминанию текста.

Таким образом, абзац – это реально существующая единица речевой практики, сознательно выделяемая автором для облегчения восприятия и усвоения его мыслей читателем. Изучая структуру абзаца по текстам различных типов и сопоставляя цепочки предложений, не принадлежащих одному абзацу, можно выявить особенности абзаца и включенных в него предложений [269, 4, 145, 210, 235, 355]: а) предложения, входящие в абзац, должны описывать одну микротему (или

несколько взаимосвязанных микротем), т.е. в составе денотативных и коннтативных значений слов, пронизывающих абзац, должны находиться общие компоненты [4, с.14]; б) предложения одного абзаца должны иметь специальные средства связи между собой; в) между предложениями одного абзаца должны участвовать определенные синтаксические связи, отражающие социально утвердившуюся логику, в соответствии с которой автор организует восприятие читателя [208, с.55; 239, с.33-34].

Для построения семиотической схемы речевого знака-абзаца необходимо отметить и такие положения: а) разные части абзаца несут различную информационную нагрузку (как и предложения) [4, 210, 235]; б) абзац в целом отражает нечто общее в ряду подобных отношений предметов действительности; в) абзац может вносить в текст определенные стилистико-экспрессивные элементы [145, с.258].

Итак, все перечисленные свойства абзаца дают нам возможность представить знаковую ситуацию абзаца так, как это показано на *рис. 3.*

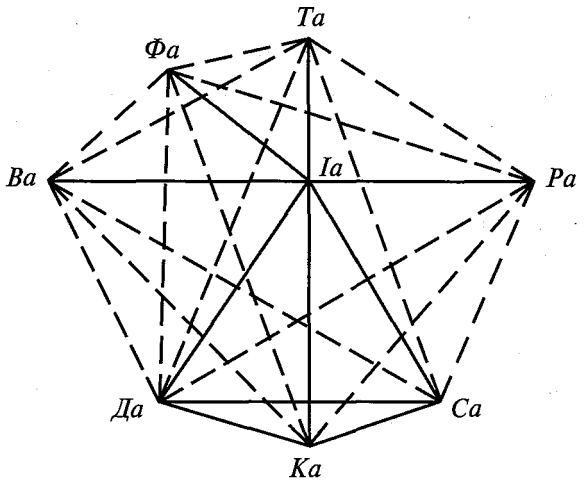


Рис. 3. Семиотическая структура знака-абзаца. *Да* – денотат абзаца; *Са* – десигнат; *Ка* – коннотат; *Ia* – цепочка из "n" имен абзаца; *Ва* – синтаксический аспект абзаца; *Ра* – информационный аспект; *Та* – тематический аспект; *Фа* – связующий межфразовый аспект абзаца.

Приведем некоторые пояснения компонентов, входящих в семантическую структуру абзаца. Если любой абзац, представляющий собой некоторый "минитекст", описывает какую-то конкретную микроситуацию окружающей действительности [99, с.17], то денотатом абзаца *Да* служит совокупность предметов этой конкретной ситуации [84, с.105; 208, с.54]. Представление о тех множествах предметов, образующих такую же микроситуацию, которая представлена в данном абзаце, является компонентом знака-абзаца – десигнат *Са.* Синтаксическое значение *Ва* абзаца – это социально значимый способ передачи отношений между предметами, входящими в описываемую абзацем микроситуацию. Как и в знаковой ситуации предложения, для абзаца значимыми являются информационный *Ра* и коннотативный *Ка* аспекты. Как видно из *рис.3,* в семиотическую структуру абзаца добавлены (по сравнению со структурой предложения) две составляющие: *Та* и

Фа. Здесь тематический аспект абзаца *Та* связан с единством темы всего текста, а связующий межфразовый аспект абзаца *Фа* отражает, в основном, наличие в абзаце специальных средств связи между входящими в абзац предложениями.

Таким образом, согласно представленной нами семиотической структуре знака-абзаца, содержанием абзаца будет являться отражение в нашем сознании отношения линейной последовательности имен абзаца к отраженной в сознании человека микроситуации как совокупности семи компонентов означаемого этого знака.

1.6. Текст как знак

Как было уже отмечено, всякий связный текст формально, но в то же время и семантически членится на абзацы, разделы, главы, параграфы и т.д. В любом его виде связный текст имеет свои, как пишет М. П. Брандес, формы организации, выработанные и кодифицированные общественной практикой [56]. По высказыванию Е. А. Реферевской, текст не является простой фиксацией речи во всех ее формах, а обладает определенными законами построения, смыслового и формального соединения и разграничения составляющих его единиц [240, с.36-37]. Известно, что любой текст отражает определенный фрагмент объективной действительности. Автором текста эта действительность представляется как совокупность огромного числа взаимосвязанных между собой предметов, явлений, свойств, процессов, отношений и т. д. [199, с.12-15; 237, с.12; 267, с.15; 336, с.213]. Такое множество взаимосвязанных объектов в литературе обозначают термином "ситуация", который также определен неоднозначно [60; 152, с. 137; 154, с.73; 178, с.446; 225, с.9; 267, с.22-31]. Мы же будем придерживаться того определения понятия "ситуация", которое дано в работе [267, с.23]: под ситуацией будем понимать определенным образом организованную или понимаемую совокупность элементов реального или воображаемого мира, преобразованную сознанием автора. Здесь нельзя не согласиться с тем, что в сознании разных людей одна и та же ситуация может восприниматься неодинаково, т.е. невполне идентичными психическими ситуациями. Причиной этому является в основном индивидуальная особенность степени восприятия человеком окружающего мира. Кроме того, это может зависеть и от позиции, с какой будет рассматривать данный индивид реальную ситуацию [51, с.81]. Но исследователи во всех этих психических ситуациях отмечают нечто общее – главное в представлении людей [178; 267, с.22-23], которое можно обнаружить, например, при чтении одного текста, путем суммирования впечатлений разных читателей. При этом можно вывести некоторую усредненную оценку, позволяющую судить о семантике текста в целом [232, с.95].

Из изложенного вытекает, что текст содержит описание некоторой "психической ситуации", которая находится под сознательным влиянием автора, в трансформированном виде. В зависимости от авторского замысла, а также от поставленной его цели и условий коммуникации в тексте происходит трансформация ситуации. Автор текста из имеющего у него арсенала составляющих и отношений психической ситуации производит выбор, включая в текст лишь некоторые, одновременно сопровождая их авторскими размышлениями, предположениями, выводами и логическими соображениями [46, с.27; 101, с.31; 208, с.54; 274, с.141]. Итак, такую трансформированную автором психическую ситуацию в литературе называют то "содержанием" [208, с.54; 261, с.20], то "темой" [308, с.11-12; 329,

с.15] или "смыслом" [49, с.37; 274, с.141-142; 311, с.15]. В дальнейшем изложении мы будем пользоваться терминами "тема" и "смысл" текста.

Текст как система должен иметь составляющие его единицы и правила их объединения в одно целое, т.е. иметь определенное внутреннее устройство. Во-первых, содержание всего текста включает содержание входящих в него абзацев. Во-вторых, основным критерием, объединяющим их в единый текст, будет служить тематический критерий. Исследования ученых, касающиеся внутренней структуры текста, показывают наличие во всем тексте некоторого объективного лексического ядра, организованного общностью предмета описания, повествования или рассуждения [209, с.27-28; 315, с.1-8]. Отсюда вывод: темой текста можно назвать отражение в сознании человека наиболее существенных составляющих данной ситуации, описанной автором в тексте. Необходимо отметить и такую особенность, что тема текста – это тот инвариант, который воспринимается одним и тем же и для автора, и для читателя (воспринимающего данный текст).

Тематическую связь в тексте называют внутренней связью. Она реализуется лексико-грамматическими средствами. Кроме нее выделяют так называемую внешнюю или формальную связь. К последней отнесем те синтаксические средства, которые связывают абзацы текста в общий синтаксический каркас текста. Такая связь пока еще мало изучена, но наличие ее считается бесспорным.

Отметим еще одну проблему, связанную с передачей информации автором и восприятием ее адресатом. В общей структуре информационного процесса важная роль отводится приемнику сообщения – адресату. Каждый адресат имеет разную степень восприятия информации, так как он обладает разной прагматикой, завсящей от разных факторов. В результате каждый читающий (адресат) извлекает из текста информацию не в полной мере по сравнению с информацией, которую вкладывал в текст автор [53, с.62; 315, с.4]. Такую особенность текста мы будем использовать при построении общей семиотической структуры знака-текста.

Проблема передачи информации автором и восприятия ее порой рассматривается в русле прагматики и связана с целевой установкой создания текста. Прагматика текста, т.е. корреляция лингвистических признаков и экстралингвистических условий, представляет собой вербальную реализацию коммуникативных установок участников общения – "... выбор языковых средств из наличного репертуара для наилучшего воздействия" [266, с.325], который во многом определяет законы текстообразования, т.е. синтаксис текста. Целью создания текста может быть простая передача всем известной информации, передача новой информации, возбуждение каких-то эмоциональных состояний адресата и т.д. Подобная особенность целого текста, существенно отличающая его от предложения и абзаца, является одной из важнейших его характеристик. На основании отмеченных свойств текста мы вполне можем утверждать, что текст обладает всеми теми признаками, которые присущи знаку. Об этом свидетельствуют и многие факты, приведенные в многочисленных научных источниках [56, с.65; 237, с.12; 331, с. 91; 333, с.19; 336, с.213; 339, с.29; 340, с.88].

Таким образом, под содержанием текста мы будем понимать отражение в сознании человека отношения линейной последовательности имен текста к отраженной в сознании ситуации как совокупности восьми компонентов означаемого этого знака. В последующих главах показано, какие знаковые составляющие текста и его единиц играют определяющую роль в случае функционирования текста в системе "человек – ЭВМ" при его формальном представлении в памяти современных компьютеров.

Итак, семиотическая структура знака-текста представлена нами как на *рис. 4.*

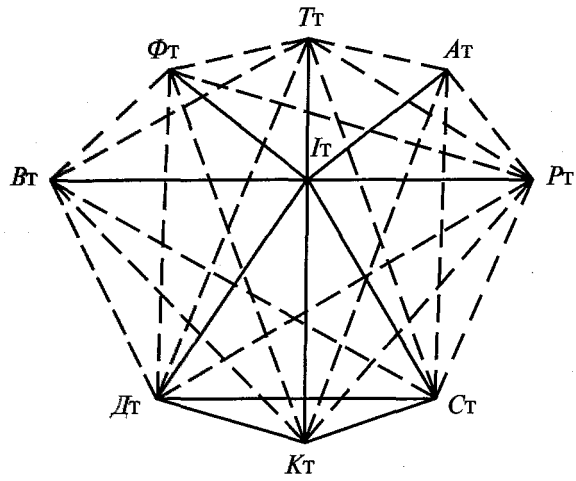


Рис.4. Семиотическая структура знака-текста. *Д*т – денотат текста; *С*т – десигнат; *К*т – коннотат; *І*т – цепочка из "n" имен текста; *В*т – синтаксический аспект текста; *Ф*т – межабзацный связующий аспект; *Т*т – тематический аспект; *Р*т – информационный аспект; *А*т – целевой аспект текста.

Глава вторая
ФОРМАЛЬНОЕ ПРЕДСТАВЛЕНИЕ СЛОВАРЯ КАЗАХСКОГО ЯЗЫКА

2.1. Индивидуальный словарь носителя языка и его роль в восприятии и понимании текста

В вероятностно-статистических закономерностях проявляются всеобщие связи явлений в природе и обществе, которые, как правило, измеряются отношением, частотой или вероятностью. Говоря о частоте лингвистических явлений, можно утверждать, что она относится к универсальным лингвистическим категориям. Здесь речь идет о том, что языку объективно присущи количественные признаки [85, с.11], так как они внутренне связаны с качественными характеристиками языка. Не оспоримым является тот факт, что, как бы разнообразны ни были естественные языки, при их функционировании наблюдаются повторяющиеся языковые явления. Наблюдая за частотой различных речевых явлений, можно судить о наличии или отсутствии статистических закономерностей в системе языка в целом или в каких-то его подсистемах. С этой целью составляются различные частотные словари слов, словоформ, словосочетаний или грамматических явлений. На материалах частотных словарей решаются многие лингвистические задачи как традиционного характера, так и задачи, относящиеся к прикладным аспектам.

Остановимся на некоторых из них. В традиционном плане частотные словари позволяют глубже исследовать лексическую организацию языка, а именно выделить лексику отдельных функциональных стилей, их общую часть, разграничить национальную лексику от заимствованной, общеупотребительную от терминологической. Частотный словарь позволяет выделить неологизмы, архаизмы и другие группы лексических единиц. Основываясь на статистических характеристиках слов, можно установить роль, место и функции этих слов в общей языковой системе.

Через количественные соотношения слов и словоформ можно говорить о лексическом богатстве и стилевых различиях языка, о способах образования новых слов. Частотные словари помогают и в сопоставительно-типологическом изучении родственных и разноструктурных языков [28, 143, с.3]. По поводу полезности частотных словарей в исследованиях традиционного плана хорошо сказано Л. Н. Засориной: "Материалы частотных словарей исключительно ценны и для собственно лингвистических исследований. Они оказывают влияние на судьбы традиционной лексикографии. Ими пользуются в решении основной проблемы общей лексикологии – выделение словарного фонда активного и периферического словаря; они полезны и при изучении вопросов стилистики, семантики и литературной нормы языка" [113, с.3-4].

Очень велика роль частотных словарей в решении прикладных задач. Особо можно отметить их значение для машинного перевода с одного языка на другой, а также при индексировании, аннотировании или реферировании научно-технических текстов. Без таких словарей невозможны автоматизация лингвистических работ,

организация эффективного информационного поиска и ряд других прикладных работ. Р. Г. Пиотровский считает, что для машинного перевода особо важно составить частотные словари наиболее часто употребляемых и информационно эффективных слов [230, с.5]. С учетом избыточности естественных языков, материалы частотных словарей могут быть использованы для рациональной организации преподавания языков, а именно при составлении словарей-минимумов, которые необходимы для изучения национальных и иностранных языков. Можно согласиться с утверждением исследователей о том, что "разумная методика обучения языку должна основываться в первую очередь на статистических характеристиках языковых явлений. При составлении учебных пособий, словарей-минимумов, терминологических справочников и словарей главным критерием отбора материала должен стать принцип частотности ..." [8, с.209]. В полной мере это справедливо и при обучении казахскому языку.

Известно, что в каждом языке выделяются функциональные стили или "подъязыки", которые обладают, по сравнению с языком в целом, ограниченным количеством лингвистических единиц и определенной спецификой их употребления. Выбирая наиболее частые и информационно насыщенные единицы языка, а также типовые контексты их употребления, можно построить такую модель языка (базовый язык), которая будет являться некоторым приближением к реальной системе подъязыка, порождающей тексты узкой тематики [231, с.21-23]. Критерием отбора высокочастотных лингвистических единиц, специфичных для определенного подъязыка, является процент покрываемости такими словами массива текстов этого подъязыка. Для более ясного представления этого понятия ("покрываемость текста") приведем примеры из статистических исследований тюркских и индоевропейских текстов. Известно, что всякий достаточно протяженный текст, независимо от его характера и языка, обладает следующим свойством: соотношения между числом "самых частых" слов и "длиной" текста, покрываемого этими словами в процентном выражении, имеют близкие значения. Иначе говоря, если по достаточно большому текстовому массиву составить частотный словарь и расположить в нем слова в порядке убывания их частот (по степени активности), то оказывается, что первые 2000 слов словаря занимают примерно 0,8 всего массива, т.е. заполняют около 80% его совокупной длины [85, с.151].

Приведем данные покрываемости текста группой наиболее частых слов по некоторым тюркским языкам. Пока еще не издан частотный словарь казахского языка в целом, но составлены частотные словари по отдельным жанровым разновидностям. В *таблице 1* приводятся данные покрываемости текстов группой наиболее частых слов частотных словарей, составленных по различным жанрам (или подъязыкам) казахского языка [1, 15, 21, 29-35, 38, 170]. Как видно из данной таблицы, различные жанры казахского языка, хотя имеют свою специфику по характеристике покрываемости длины текста высокочастотными словами, все же они колеблятся между некоторой средней величиной. Увеличением процента покрываемости отличается поэтический текст и текст сказок. Средняя величина покрываемости для рассматриваемых шести текстов такова: 1000 наиболее частых слов покрывают 76,58%, 2000 слов – 85,80% и 3000 слов – 88,70% массива текста. В седьмом тексте, т.е. в 20-томном собрании сочинений М. Ауэзова, смешаны тексты разных жанров, несмотря на это, проценты покрываемости текста группой высокочастотных слов отражают картину, свойственную различным казахским текстам.

Можно было бы привести данные и по другим группам тюркских языков, например, согласно данным "Словаря наиболее употребительных слов современного узбекского языка" [147] 1227 лексем из высокочастотной зоны (это 15,2 % длины словника) покрывает 76% всего обследованного текста, т.е. 76811 словоупотреблений.

Если сравнить с данными для казахского языка, где 1000 наиболее употребительных слов покрывает 76,58% длины текста, то здесь больших расхождений не наблюдается.

Исходя из приведенных данных, напрашивается естественный вопрос: можно ли сделать вывод о том, что человек, изучающий казахский или узбекский языки и знающий эти 1000 (или 1227) наиболее употребительных слов, сможет понимать около 80% текста современной казахской (или узбекской) художественной прозы? Видимо это не совсем достаточно, так как в их числе немало многозначных лексем, а некоторые из них входят как компоненты устойчивых словосочетаний и фразеологических единиц. Поэтому, пишет И. А. Киссен, для понимания 80% слов текстов художественной прозы необходимо хорошо знать примерно 2–3 тысячи слов языка, а остальные 20% лексем текста можно будет понять по догадке или на основе знания правил словообразования [147, с.6-7].

Таблица 1. Покрываемость текста группой наиболее частых слов

№ п/п	Тексты	Объем текста	Длина словника	Ранговый интервал слов и % их покрываемости					
				1-50	1-100	1-500	1-1000	1-2000	1-3000
1	М. Ауэзов "Абай жолы"	465373	17188	30,5	39,2	62,0	71,6	79,9	84,0
2	М. Ауэзов "Өскен өркен"	90929	7674	31,0	41,3	68,0	70,2	89,2	93,2
3	Абай Поэзия	46847	6017	33,1	45,0	69,2	89,7	88,9	93,4
4	Газетные тексты	146764	12423	23,0	31,7	59,7	72,8	83,9	88,9
5	Детская литература	98040	10076	26,8	35,4	58,4	69,1	79,5	86,1
6	Казахские сказки	98058	5840	41,5	51,8	76,6	86,0	93,4	96,6
7	М. Ауэзов. 20- томн. пр.	1706195	29483	31,3	41,4	66,3	76,4	85,7	90,1

По другим группам языков данные покрываемости в основном сходятся. Например, по словарю М.-Н. О. Османова "Частотный словарь Унсури" (Унсури – персидско-таджикский поэт XI века, текст "Дивана") 1863 более частых слова (38,62% длины словника) покрывает 92% текста "Дивана", общая протяженность которого составляет 46472 словоупотребления [221, с.10].

Для английского языка (по Торндайку) покрываемость 2000 слов из высокочастотной зоны составляет 78%, чешского языка (по Тешетыловой) – 75%, русского языка (по "Частотному словарю русского языка" под редакцией Л. Н. Засориной) – 76% [304].

Как замечает Р. М. Фрумкина, частотный словарь в силу ряда методических соображений не может быть использован в качестве словаря-минимума, но он является необходимой и единственно надежной основой для его составления [296, с.20]. Как показывают исследователи [85; 295–297 и др.], критерий частоты необходимо дополнить не только критерием "обиходности", но также рядом лингвистических и дидактических критериев.

Интересен специальный эксперимент, проведенный Р. М. Фрумкиной для проверки на эффективность словаря Э. А. Штейнфельдт. Была поставлена такая задача: выяснить, в какой мере данные, полученные по текстам ограниченного объема и жанра, будут верны для других русских литературных текстов, относящихся к иным жанрам и периодам. С этой целью был подсчитан процент покрываемости текста наиболее частыми словами для 60 различных текстов длиной в 1000 слов. Эксперимент показал, что первые 1300 и тем более первые 2000 слов дают вполне удовлетворительный средний процент покрываемости, колеблясь от средней величины от 0,68 до 0,81 [296, с.21-22].

Покрываемость различных текстов высокочастотными словоформами для отдельных тюркских групп языков подробно изложена в книге К. Б. Бектаева "Статистико-информационная типология тюркского текста" [28, с.37-39]. Отметим только, что покрываемостm казахских текстов имеет различия, например, от индоевропейских языков. Если в последнем 100–150 словоформ покрывают 50% обследуемого текста, то в казахском подъязыке публицистики и в художественном тексте (роман "Абай жолы") такой процент покрываемости текста соответствует 700–800 словоформам [15, с.553; 91, с.51]. К. Б. Бектаев [28] приводит такие сравнительные данные по покрываемости текстов: в высокочастотной зоне покрываемость 1000 первых казахских словоформ колеблется от 47 до 60%, а для индоевропейских языков – от 62 до 85%. Когда единицей словаря служит не словоформа, а слово, то на 1000 слов высокочастотной зоны словаря покрываемость по казахским текстам варьирует от 63 до 86%, а по текстам индоевропейских языков – от 64 до 89% .

Такое различие в процентах покрываемости казахских словоформ К. Б. Бектаев объясняет тем, что "... в тюркских языках агглютинативная конструкция по своим функциям часто идентична не столько словоформе, сколько словосочетанию в индоевропейских языках" [28, с.39]. Далее, автор расскрывает причину такого несоответствия тем, что в языках агглютинативного строя слова имеют более разнообразные виды формообразования по сравнению с языками флективного строя. Действительно, в казахском языке только у существительных количество формообразующих аффиксов достигает 500, а большинство казахских глаголов имеют до 1000 различных форм [16, 36].

Таким образом, сравнительно небольшая группа наиболее частых слов в сумме составляет огромное число словоупотреблений. Последнее дает возможность, рассматривая незначительную часть инвентаря языка, зафиксированную в частотном словаре, делать выводы о большей части единиц речи.

Подводя итог (вслед за И. А. Киссеном), можно сделать следующий вывод об объеме индивидиуального словаря носителя казахского языка: для понимания 80% казахского текста достаточно хорошо знать значения казахских слов в пределах от 2 тыс. до 3 тыс. слов из высокочастотной зоны частотного словаря, составленного по различным жанрам этого языка. Оставшиеся 20% значений лексем можно узнать по отдельным отраслевым терминологическим словарям или по словарю новых слов. Не исключено, что значения этих слов можно будет понять из соответствующего контекста и по догадке. Последнее, в основном, зависит от уровня знания данного предмета (о котором идет речь в тексте) индивидуальным носителем языка.

В нашей работе будет рассматриваться вопрос об установлении основных принципов формализации содержания казахского текста, в частности статического содержания. Решение такой сложной задачи (особенно для тюркоязычных текстов) основывается нами на принципах частотности лингвистических единиц текста с опорой на высокочастотные зоны словаря и на характеристику равномерной покрываемости текстов. Так как задача формализации содержания казахского текста будет решаться в целях использования современных компьютеров, то возникает

ряд вопросов. В частности, какой минимальный объем словаря казахского языка необходимо записать в память компьютера? По каким областям знаний (подъязыкам) лучше всего выбрать экспериментальные тексты? Какие слова считать "опорными", "главными", вокруг которых сконцентрировано содержание текста? Эти и другие вопросы, касающиеся раскрытия принципов формальной структуры содержания казахских текстов, будут рассмотрены нами детальнее в следующих разделах.

Итак, для более полного охвата различных областей знания были взяты тексты из художественной литературы, публицистики (газетные статьи на одну тематику) и научно-популярной литературы, т.е. эксперименту подвергнуты тексты, относящиеся к трем функционально-смысловым типам речи, таким, как описание, повествование и рассуждение.

2.2. Лексико-морфологическая характеристика компьютерного словаря казахского языка

Для того чтобы построить лингвистическую модель, которая дала бы возможность порождать тексты заранее заданного содержания, необходимо выявить такие статические и динамические составляющие, из которых состоят тексты рассматриваемых типов. Можно полагать, что существуют определенные правила развертывания текста из заранее заданного содержания в виде последовательности некоторых предложений. Как мы уже отмечали ранее, особенность порождения текстов заключается в определенном соотношении детерминированных и вероятностных правил, действующих в данном процессе. По этой причине в лингвистической модели порождения текста должны быть учтены лингвистические единицы, специфичные для различных типов текстов. Для выявления таких правил и единиц содержания были рассмотрены три типа казахских текстов: художественный, публицистический и научно-популярный.

На первом (предкомпьютерном) этапе в указанных текстах к каждой словоформе нами были проставлены индексы, обозначающие класс слов казахского языка. Далее каждая словоформа текста была преобразована в лексему путем отделения аффиксов от слова (лемматизация). На следующем этапе экспериментальные тексты были введены в память компьютера с указанием номеров страниц для книжных текстов и номеров выпуска для газетных статей. Дополнительно были указаны порядковые номера абзацев для каждого текста.

Далее следовал этап компьютерной обработки текстов с помощью специально составленных программ на языке ТУРБА-ПАСКАЛЬ. В процессе такой обработки были получены:

1. Алфавитно-частотный словарь-словоуказатель для каждого типа текстов.

В словаре указаны порядковый номер лексемы (колонка 1); F1 – абсолютная частота лексемы в исследуемом тексте (2); F2 – количество текстов, в которых встретилась данная лексема (3); лексема с соответствущим индексом, указывающем на класс слова (4); Пij-kl, где П – тип признак текста (П – публицистический текст, Х – художественный текст, Н – научно-популярный текст), а индексы (ij-kl): а) ij – порядковый номер текста; б) kl – порядковый номер абзаца, в которых встретилась лексема (5).

2. Алфавитно-частотный словарь лексем в целом по каждому стилю с указанием абсолютной частоты слова и количества текстов данного стиля.

3. Алфавитно-частотный словарь для отдельно взятого номера текстов (ij) для каждого типа текстов (П – признак типа текстов).

4. Частотно-алфавитный словарь по отдельно рассматриваемым типам текстов (для каждого стиля). В этом словаре лексемы расположены по порядку убывания их абсолютных частот. Кроме этого, в словаре даются накопленная абсолютная и относительные частоты для каждой лексемы.

5. Обратный алфавитно-частотный словарь для каждого стиля. В словаре лексемы расположены строго по алфавиту, начиная с конечных букв в слове. Здесь также указываются абсолютная частота (F1), ее абсолютная накопленная (f2) и относительная накопленная (f3) частоты лексемы.

Фрагменты этих словарей представлены в "Приложении".

На следующем этапе эти словари были исследованы с точки зрения употребительности в них различных классов слов. Результаты исследования по рассматриваемым трем стилям показаны в *таблице 2.* В ней также приводятся данные по совокупности всех трех типов текстов.

Рассмотрим подробнее статистические данные по трем стилям казахского языка и единому словарю, который получен в результате слияния текстов трех рассматриваемых стилей. Данные *табл. 2* подтверждают известные теории о специфичности стилей казахского языка. Так, существительные в публицистике и научно-популяной литературе употребляются гораздо чаще, чем в художественной литературе. Относительно глагольных слов казахского языка можно отметить обратное, т.е. глаголы здесь употребляются на 13−14% чаще, чем в текстах двух других стилей.

Эти факты вполне объяснимы: в художественных текстах идет некоторое повествование, развитие действий, выражаемых глаголами. Для текстов научных и публицистических характерны рассуждение, объяснение, которые чаще выражаются существительными. Однако, необходимость напоминания о небольшом числе главных действующих лиц в художественном тексте приводит к возрастанию в них числа имен собственных и местоимений.

Прилагательные − основные единицы языка, которые помогают активно рассуждать и объяснять. Поэтому их число в текстах научно-популярных и публицистических больше, чем в текстах художественных.

Для классов слов, относящихся к наречиям и служебным словам, характерно, что эти части речи во всех трех типах текстов имеют почти одинаковую частоту.

Вполне объясним и факт отсутствия междометий и подражательных слов в публицистических и научно-популярных стилях.

В связи со спецификой тематики публицистического текста, выбранного для исследования ("Шестилетние идут в школу") замечается некоторое преобладание числительных, чем в остальных рассматриваемых текстах.

В едином словаре, полученном в результате объединения данных всех трех словарей, можно отметить некоторые особенности, свойственные казахскому языку. В частности, соотношение количества глаголов и существтительных в едином словаре составляет (0,2556/0,4099)=0,62, что ближе к тексту научно-популярному, объединяещему черты художественного и научного текста. В научно-популярном стиле указанное соотношение равно (0,2311/0,4354)=0,55, в то время как в художественном и публицистическом соотношение глаголов и прилагательных составляет соответственно (0,3575/0,3103)=1,1 и (0,2199/0,4430)=0,49.

Почти совпадают соотношения прилагательных и существительных в едином словаре (0,29) и в текстах трех других стилей (соответственно 0,283; 0,297; 0,294).

Отношения наречия к глаголу в едином словаре − 0,184, а в художественном тексте − 0,115, в публицистическом − 0,191 и в научно-популярном − 0,235, т. е. они почти совпадают, колеблясь около 0,2. Некоторое отклонение от этого значения отмечается в художественном стиле текстов (0,1).

Таблица 2. Употребительность классов слов в трех жанровых разновидностях казахского языка

№ п/п.	Класс слов	Тип текста и частоты								
		художественный		публицисти-ческий		научно-популярный		единый словарь по 3-м стилям		
		абсол. F	относит. f	абсол. F	относ. f	абсол. F	относ. f	Кол.сл. n	абсол. F	относ. f
1	Существительное	4235	0,3103	9621	0,4430	10991	0,4354	2514	24847	0,4099
2	Имя собственное	451	0,0330	598	0,0275	301	0,0119	470	1350	0,0223
3	Местоимение	917	0,0672	573	0,0264	634	0,0251	101	2124	0,0350
4	Числительное	294	0,0215	777	0,0358	564	0,0223	68	1635	0,0270
5	Глагол	4880	0,3575	4776	0,2199	5835	0,2311	1939	15491	0,2556
6	Прилагательное	1200	0,0879	2866	0,1320	3233	0,1281	1507	7299	0,1204
7	Наречие	561	0,0411	913	0,0420	1372	0,0543	426	2846	0,0470
8	Служебные слова	754	0,0552	1303	0,0600	1944	0,0770	85	4001	0,0660
9	Междометие	48	0,0035	4	0,00001	—	—	31	52	0,0001
10	Подражательные слова	72	0,0053		0,00000	1	0,0000	54	74	0,0012
11	Модальные слова	238	0,0174	288	0,01330	370	0,0147	24	896	0,0155
	Всего	13650	1,0000	21720	1,0000	25245	1,0000	7219	60615	1.0000

2.3. Автоматический словарь казахского языка (АСКЯ)

2.3.1. Формальное описание словарной статьи

Известно, что словарь модели лингвистической информации о слове строится по машинным основам изменяющихся слов. Такие основы для русского языка, например, выделяются путем отсечения от словоформ окончаний и суффиксов в соответствии с разработанными для данной модели типами формообразований различных классов слов [123]. Для казахского же языка, который по своей типологии и морфологической структуре относится к группе агглютинативных языков, нет особой необходимости выделять такие машинные основы путем формального отсечения, так как сам строй данного языка благоприятствует решению этого вопроса. Разработка этой проблемы помогла бы поставить на повестку дня исследовательской практики более важную проблему – создание теории формальной грамматики казахского языка, которая дала бы ощутимые результаты в компьютерной обработке современных письменных текстов. Известно, что имеющиеся академические грамматики казахского языка преследуют во многих случаях иную цель, поэтому создание формальной грамматики казахского языка, понятной компьютерным языкам, должно быть, на наш взгляд, одной из неотложных задач ученых-казаховедов. При этом, как нам кажется, узловой единицей, стержнем вокруг которого будет группироваться языковая информация, должно стать **слово,** как наиболее естественная единица языка. Такая постановка вопроса обосновывается, по нашему мнению, в силу того, что слово можно структурировать и в глубину – на морфемы, фонемы, слоги, буквы, значения, и во вне – как член словоизменительных, словообразовательных, синтаксических, грамматических, ассоциативных и концептуальных рядов [139]. Кроме того, в каждом слове амальгамирована его история, которая связана с историей народа и его культуры. Поэтому-то "слово" является объектом изучения и лексикологии, и семасиологии, и морфологии, и синтаксиса словосочетаний, а также объектом стилистики и других областей языкознания. Все это свидетельствует о том, что формальное описание словарной статьи "Автоматического словаря казахского языка" должно быть организовано так, чтобы всю соответствующую информацию, сконцентрированную вокруг данного слова, в любой момент можно было бы извлечь, а амальгаму – соответствующим образом расшифровать. В основе способа хранения такой информации должно лежать своеобразное "досье" на каждое слово, представленное как подробно структурированная область базы данных, содержащая какое-то число потенциальных сегментов, которые отражают всю систему фонетических, морфологических, синтаксических, семантических, семасиологических, стилистических, историко-этимологических, диалектологических, контекстуальных и т.п. параметров, т.е. всех возможных сведений о слове.

Такие автоматические словари, как АСКЯ и формальная грамматика, могут быть жизнеспособными и эффективными, если они опираются на общую филологическую традицию и культуру, на глубокое изучение казахского языка и учет информации о нем во всех формах его существования. Создание модели лингвистической и энциклопедической информации о казахском слове позволит практически осуществить конкретное моделирование лексико-семантической, словообразовательной, формообразавательной и других структур казахского языка.

Мы разработали лишь обобщенную схему такой области, содержащую некоторую лингвистическую информацию о казахском слове. Составлением же принципиальной схемы автоматического формирования "досье" и ее реализацией на современных компьютерах должны заняться специалисты по созданию автоматизированной базы лингвистических данных.

2.3.2. Структура автоматического словаря казахского языка

С учетом того, что "слово" любого языка по своей природе является сложным лексико-грамматическим единством, необходимо проводить разделение слов на определенные структурные группы или типы, согласно выбранному для каждого случая принципу классификации. Наш принцип создания автоматического словаря казахского языка (АСКЯ) преследует частный случай и не претендует быть универсальным словарем, охватывающим все случаи исследовательской практики. В основе нашего АСКЯ лежит алфавитный словарь слов-основ казахского языка, где к каждому слову представлена лингвистическая информация – словарная статья, которая представляет собой отдельные зоны (области памяти), в которых записана условно принятая нами кодовая информация о некоторых морфологических, фонетических, семантических, синтаксических и других признаках (параметрах) рассматриваемого слова. Слово-основу, в данном случае, следует понимать как звуковой комплекс с определенным значением, представляющий собой корневые и производные (аффиксальные) слова, относящиеся к определенному классу слов казахского языка. Слово-основа и есть основная лингвистическая единица АСКЯ, к которой приписана соответствующая словарная статья, состоящая из лингвистической (и энциклопедической) информации, записываемой в виде формально принятых нами кодов в указанном номере зон. Количество зон для каждой единицы словаря будет зависеть от максимального количества "досье" к данному слову. Причем каждая зона с одним и тем же номером должна иметь информацию одного и того же типа (содержания). Например, зона 1 сообщает информацию о принадлежности рассматриваемого слова к определенному классу слов: существительное, глагол, прилагательное и т.д. При этом некоторые зоны для отдельных слов "временно" могут иметь "нулевую" (пустую) информацию до тех пор, пока его содержание окончательно не выяснится. Это даст возможность постоянно редактировать и расширять объем "досье" к слову.

Схематически **АСКЯ** имеет следующий вид:

Слово-основа	Зона 1	Зона 2	Зона 3	...	Зона i	...	Зона n
← ..... Словарная статья ..... →							

В случае, когда содержание информации для данной зоны "компактное", в смысле занимаемого объема памяти, то его можно поместить в соответствующем выделенном номере зоны. Если же такая информация о слове требует дополнительного объема памяти (чем выделено) или же она еще не существует в готовом виде, то такая зона должна иметь соответствующую команду, отсылающую на выполнение задания по поиску или созданию такой информации.

Например, при синтезировании всевозможных словоформ для данного слова-основы или же при снабжении данного слова контекстными примерами из разных

иллюстрационно-текстовых источников и др. Этот факт говорит о том, что "досье" к слову в автоматическом словаре может быть не всегда доступным потребителю непосредственно, т.е. оно не всегда известно в готовом виде. В таких случаях должен быть указан путь достижения цели (алгоритм). Иначе говоря, в этом случае необходимо провести соответствующее исследование по заранее указанному алгоритму и формировать требуемую информацию для данной зоны.

Таким образом, АСКЯ должен как бы обладать "пониманием" известных языковых законов, т. е. по возможности в АСКЯ должны быть смоделированы синхронные и диахронные языковые процессы, которые представлены в динамике, позволяющие функционировать отношения единиц, уровней, подсистем и т.д. Следовательно, автоматический словарь казахского языка **(АСКЯ)** должен существовать в компьютере не как готовая книга-словарь, а как научная лаборатория, банк научных знаний об объекте (в данном случае, о казахском слове). В перспективе АСКЯ должен стать не только хранилищем знаний о языке, но и представлять собой "разумный", специализированный банк знаний, способный предложить решение или нацелить исследователя на определенное решение.

По нашей версии, порядок расположения информации, т.е. информация в соответствующих номерах зон, может быть произвольным, но стандартным для всех единиц словаря.

Зона 1. *Слово из автоматического словаря принадлежит к одному из ниже перечисленных классов слов:*

№ п/п	Класс слова	Сокращен.	Код
1	Имя существительное *(зат есім)*	ЗТ	01ЗТ
2	Глагол *(етістік)*	ЕТ	01ЕТ
3	Имя прилагательное *(сын есім)*	СН	01СН
4	Местоимение *(есімдік)*	ЕС	01ЕС
5	Имя числительное *(сан есім)*	СА	01СА
6	Наречие *(үстеу)*	УС	01УС
7	Служебные слова *(шылау сөздер)*	ШЛ	01ШЛ
8	Подражательные слова *(еліктеуіш)*	ЕЛ	01ЕЛ
9	Междометие *(одағай)*	ОД	01ОД

Зона 2. *Слово из автоматического словаря относится к одному или нескольким рассматриваемым подъязыкам:*

№ п/п	П о д ъ я з ы к и	Сокращен.	Код
1	Художественный	ХД	02ХД
2	Публицистический	ПБ	02ПБ
3	Научно-популярный	НП	02НП
4	Художественный и публицистический	ХП	02ХП
5	Художественный и научно-попул.	ХН	02ХН
6	Публицистический и научно-попул.	ПН	02ПН
7	Худ-й, публ-й и научно-популярный	Х-	02Х-

В третьей главе нами проведено исследование по семантической классификации знаменательных классов слов казахского языка (существительное, глагол, прилагательное, наречие и местоимение). В *табл.* 3–10 указанной главы показаны содержания (примеры) семантических классов и подклассов указанных частей речи, а также условные компьютерные коды соответствующих подклассов. Данные этих таблиц позволят заполнить "**зону 5**" словарной статьи для соответствующей лексемы АСКЯ.

Содержание словарной статьи в АСКЯ в начальных номерах зон, по предлагаемой нами версии, должно выглядить так, как это изложено ниже:

Зона 3. *Слово из автоматического словаря по способу образования относится к одному из ниже перечисленных:*

№ п/п	Способ образования	Сокращен.	Код
1	Корневое *(түбір сөз)*	ТС	03ТС
2	Аффиксальное *(аффиксті)*	АФ	03АФ
3	Слитное *(біріккен сөз)*	БС	03БС
4	Парное *(қос сөз)*	ЕЕ	03ЕЕ
5	Повторное *(қайталама сөз)*	ҚҚ	03ҚҚ
6	Сокращенное *(қысқарған сөз)*	СҚ	03СҚ

Зона 4. *Слоговой состав слова из автоматического словаря включает:*

№ п/п	Слоговой состав слова	Сокращен.	Код
1	Односложное *(бір буынды)*	ББ	04ББ
2	Двусложное *(екі буынды)*	ЕБ	04ЕБ
3	Многосложное *(көп буынды)*	КБ	04КБ

Зона 5. *Слово из автоматического словаря относится к соответствующему семантическому классу.*

Таким образом, все изложенное по первым пяти зонам на примере парного слова *"әке-шеше"* в АСКЯ будет иметь следующий вид:

әке-шеше	013Т	02ХД	03ЕЕ	0404	053Т059

слово-основа ..... зоны с 1 по 5 .....

Ввиду того, что порядок расположения информации в словарной статье может быть произвольным, то далее, по усмотрению создателя АСКЯ, можно продолжить все предполагаемые лингвистические и энциклопедические информации в их кодовых значениях и помещать в предназначенных для них зонах словарной статьи.

Раскодирование информации в понятном для лингвиста (или для любого интересующего словарем) виде и поиск необходимой инфомации в словарной базе и др., как мы уже отмечали, – дело специалистов по созданию компьютерного АСКЯ.

2.4. Синтез казахских словоформ-существительных для АСКЯ

В АСКЯ обязательно должны быть даны все аффиксальные словоизмения, т.е. всевозможные словоформы от рассматриваемого слово-основы. С этой целью можно было бы составить алфавитный словарь словоформ по разным текстам и выбрать по ним все словоформы, соответствующие данной слово-основе. Данная процедура, во-первых, не гарантирует охват всевозможных вариантов формообразования и, во-вторых, постоянное их присутствие в АСКЯ заняло бы достаточно большой объем памяти в словарной базе. Исходя из этого, для основных классов слов казахского языка (существительное, глагол, прилагательное) мы воспользовались грамматическими правилами формирования словоформ и разработали алгоритм их порождения, т.е. провели синтез казахских словоформ по графическим формальным признакам.

Агглютинативный строй казахского языка намного облегчает задачу формирования словоформ по известной основе слова, так как любое словоизменение образуется посредством последовательного присоединения к основе слова соответствующих аффиксов, а также путем присоединения в виде цепочек одних аффиксов к другим.

Допустим, что рассматриваемое в АСКЯ слово-основа относится к определенному классу слов, тогда в зоне 6 необходимо формировать всевозможные словоформы, т.е. перечень словоизменений (грамматические значения) с одним лексическим значением.

Рассмотрим случай, когда слово-основа относится к классу существительных. Как известно, существительное в казахском языке, как часть речи, объединяет категорию знаменательных слов-имен, которые обладают общими семантическими и грамматическими признаками [278, с.130]. Кроме того, данная часть речи является одной из многочисленных и сложных по структуре. Именно за счет существительных пополняется словарный состав любого языка, в том числе, и казахского.

Морфологическая структура аффиксов существительных и их возможные сочетания были исследованы в количественном отношении и ранее [16, с.49-87]. По статистическим данным в разных стилях казахского языка, в среднем, около 40% всех слов и 45% всех словоформ относятся только к существительным. Это, видимо, объясняется не только тем, что им присущи различные наименования предметов и явлений, но и их функциями в предложении, в котором они выступают в роли подлежащего, определения, дополнения, обстоятельства и именного сказуемого и т.д.

Имени существительному также присущи характерные только для него аффиксы словообразования и словоизменения. Мы здесь будем рассматривать только словоизменительные аффиксы существительных, порождающие словоформы с одним лексическим значением.

Известно, что слово-существительное, вступая в семантико-грамматическую связь с другими словами предложения, изменяется по числам, принимая или не принимая аффиксы множественности, по падежам, принимая аффиксы простого или притяжательного склонения, и наконец, по лицам, принимая местоименные аффиксы. Последние, т.е. местоименные аффиксы (или сказуемости), присоединяются к существительным только со значениями специальности человека или родственных отношений между людми. Этот факт учитывается и в нашем исследовании.

А. Ахабаев и К. Б. Бектаев исследовали теоретически возможные формы-варианты присоединения аффиксов к основе существительных, а также частоту их встречаемости в газетных и художественных (роман "Абай жолы") текстах казахского языка. Ими были приняты следующие условные обозначения: Р – основа слова (корневая и производная); К аффиксы множественности; С – аффиксы простого склонения; Т аффиксы притяжательности; Ж – местоименные аффиксы (или аффиксы сказуемости). Согласно данным исследования, обозначенные четыре разновидности аффиксов, присоединяясь к основе слова и в рознь, и сочетаясь в определенной последовательности, теоретически могут встречаться в 496 формах-вариантах. Здесь, конечно, учитывается количественное разнообразие внутри каждой из четырех разновидностей (К, С, Т, Ж). Из возможных 496 форм-вариантов в газетных текстах были обнаружены только 225 видов (45%), в художественном тексте – 276 видов (56%). В двух стилях слово-основа Р, т.е. основа с нулевым аффиксом, в среднем составляет 40,54% всех существительных, а оставшиеся 59,46% относятся к существительным со словоизменительными аффиксами.

Какие же последовательности аффиксов присущи существительным казахского языка?

Действительно, и наши исследования, и исследования выше названных авторов подтверждают следующий возможный порядок присоединения аффиксов к основе-слову:

I. Р-С, Р-К-С, Р-К-Т-С, Р-Т-С, Р-Т-К-С;
II. Р-Т, Р-К-Т;
III. Р-К, Р-Т-К, Р-К-Ж-К;
IV. Р-Ж, Р-Т-Ж, Р-Т-С-Ж, Р-С-Ж, Р-К-Т-Ж, Р-К-Ж.

Как видно из этой записи, аффиксы К, Т, С, Ж могут присоединяться к основе Р как непосредственно (в рознь), так и сочетаясь в определенном порядке. В данном случае, мы специально сгруппировали их в четыре подгруппы так, чтобы в каждом случае замыкал цепочку один из видов словоизменительных аффиксов. Такой порядок расположения подгрупп соответствует занимаемому месту по частоте встречаемости этих подгрупп среди всех существительных со словоизменительными аффиксами. Например, на первом месте стоит подгруппа, когда замыкающим цепочку аффиксов является простое склонение – С, которое в суммарном подсчете форм составляет более 70% (в указанных стилях – 73 и 78%) всех словоформ, далее, на II месте – Т (21 и 17%), на III месте – К (7 и 4%) и на последнем IV месте – Ж (0,01 и 0,3%). Таким образом, при словоизменительной процедуре существительным казахского языка более свойственны в конце словоформы аффиксы простого склонения.

Наше исследование преследует несколько иную цель, а именно для каждой единицы АСКЯ, т.е. для корневых или производных слов-основ, принадлежащих к классу существительных, формальным путем воссоздать всевозможные аффиксальные словоизменения – словоформы. С этой целью мы за основу своего исследования взяли известное в грамматике правило, которое гласит, что для единственного числа слов в зависимости от последнего слога и звука слова-основы простое склонение имен имеет пять фонетических вариантов.

В целях удобства составления алгоритма задачи в фонетических вариантах простого склонения нами сделана незначительная перегруппировка конечных звуков (букв) основы. Например, второй фонетический вариант относится к склонению имен, оканчивающихся на гласные ***у, и,*** щелевые сонорные ***р, л,*** среднеязычный ***й***

и неслоговой ***у***. В составленных нами пяти списках аффиксов щелевой сонорный ***л*** перенесен в пятый фонетический вариант, который касается склонения имен, оканчивающихся только на звонкие согласные ***ж*** и ***з***. Такая перестановка не согласуется с общепринятой в единственном случае, а именно, когда основа, оканчивающаяся на звонкие согласные ***ж, з,*** склоняется в творительно-соединительном падеже, принимая падежные окончания ***-бен/-бенен***. Для основ с конечной буквой (звуком) ***л*** в этом случае необходимы падежные окончания ***-мен/-менен,*** *(ел-мен, ел-менен, бал-мен, бал-менен).* Это в обязательном порядке должно учитываться в алгоритме задачи синтеза именной словоформы.

Известно, что аффиксы сказуемости присоединяются к существительным, когда они означают понятия, связанные с людьми и их деятельностью, профессией и т.д. По этой причине мы такие существительные решили рассматривать отдельно и списки с обязательным участием аффиксов сказуемости также не смешывать с ранее рассмотренными двумя случаями. При составлении таких списков нам пришлось объединить два из пяти фонетических вариантов при простом склонении имен. Точнее, склонение имен, оканчивающихся на сонорные носовые ***м, н, ң,*** и склонение имен с конечными звонкими согласными ***ж, з,*** а также щелевой сонорный ***л*** мы решили сгруппировать в один тип. Естественно, при этом не обошлось без исключений. Когда именная основа в конце имеет звук ***ж*** или ***з,*** то к основе вместо аффикса *-мын* следует приписать ***-бын,*** а когда конечная буква основы – ***л,*** то вместо ***-бысыз, -бысыздар, -бысың, -бысыңдар*** нужно пользоваться аффиксами ***-мысыз, -мысыздар, -мысың, -мысыңдар,*** соответственно их заменив. Эти исключения имеют силу и для твердых и мягких конечных слогов основы.

Мы в своем эксперименте по составлению списка аффиксов для синтеза словоформ от корневых и производных основ существительных рассматривали следующие четыре случая:

1) основа оканчивается на мягкий слог (и звук) и среди присоединяемых аффиксов притсутствие аффиксов сказуемости необязательно;

2) основа оканчивается на мягкий слог (и звук) и среди присоединяемых аффиксов участие аффиксов сказуемости обязательно;

3) основа оканчивается на твердый слог (и звук) и среди присоединяемых аффиксов притсутствие аффиксов сказуемости необязательно;

4) основа оканчивается на твердый слог (и звук) и среди присоединяемых аффиксов участие аффиксов сказуемости обязательно.

Для случаев 1) и 3), т.е. когда среди присоединяемых аффиксов притсутствие аффиксов сказуемости необязательно, на основе пяти фонетических вариантов простого склонения нами составлены **по пять типовых списков аффиксов**, присоединение которых к каждой основе порождает именную словоформу. А для случаев 2) и 4), т.е. когда среди присоединяемых аффиксов участие аффиксов сказуемости обязательно, было составлено уже **по четыре типовых списков аффиксов**, также позволяющих создать соответствующие именные словоформы.

Для удобства дальнейшего изложения составленные нами списки аффиксов будем именовать как *список 1, список 2, список 3* и *список 4.* Дадим некоторое разъяснение:

1. *Список 1* это список аффиксов, который должен использоваться для порождения словоформ от корневых и производных основ существительных с конечным мягким слогом (звуком) в реестре АСКЯ. При этом, в сочетании аффиксов могут не участвовать аффиксы сказуемости. Список задан в пяти

колонках. В первой колонке (тип 1) дан перечень аффиксов, соответственно типу слов-основ и конечных звуков (букв), т.е. основы, оканчивающиеся на гласные, кроме ***у, и,*** как, например, *әке, келі, мекеме* и т.д. В остальных четырех колонках списка (тип 2, тип 3, тип 4, тип 5) даны фонетичские варианты (с незначительными перегруппировками, о которых было отмечео выше). В соседних колонках аффиксы и их сочетания, в большинстве случаев, могут повторяться, но среди них есть иногда и специфичные для данной колонки фонетические варианты. Запись списков аффиксов в пяти колонках осуществлена специально с целью избежать многократных проверок конечных слогов и звуков (букв) у именной основы в АСКЯ, чем и достигается экономия во времени. О некоторых исключениях, требующих минимальных проверок, нами было уже отмечено. В данном случае основной проверкой является отнесение реестрового слова из АСКЯ к одному из четырех списков (со *списка 1* по *4*) и согласно конечному слогу (звуку) именной основы выбор типа колонки по сответствующим фонетическим вариантам.

2. *Список 2* отличается от с*писка 1* тем, что он предназначен для именных основ, которые по своему значению относятся к понятиям, связанным с людьми, т.е. с их родственными отношениями, деятельностью и профессией и т. д. Поэтому в *списке 2* в четырех колонках даны специальные для таких случаев аффиксы сказуемости или так называемые личные окончания (в сочетании с другими видами аффиксов) для существительных с конечным мягким слогом (звуком) в реестре АСКЯ. Список задан в четырех колонках.

3. *Список 3* отличается от предыдущего *списка 1* только тем, что он предназначен для именных основ с твердым конечным слогом (звуком) из АСКЯ.

4. С*писок 4* в отличие от *2* предназначен для именных основ с конечным твердым слогом (звуком) в реестре АСКЯ. В остальном все перечисленные функции *списка 2* присущи и *списку 4.*

Методика составления всех четырех перечисленных списков почти одинакова, а именно по первому типу, т.е. когда конечный звук (буква) является гласным (кроме *у, и*), составляется перечень всевозможных вариантов аффиксаций типовых слов, взятых для примера, и полученный словарь аффиксов сортируется по алфавиту. Остальные колонки, т.е. другие фонетические варианты *списка,* заполняются анологично первой на основе взятых для примера типовых слов. Составленные по такой методике в с*писках 1* и *3* всевозможные случаи аффиксации именной основы по первому фонетическому варианту (1 колонка) в нашем эксперименте оказались равны 130, а в остальных колонках (фонетических вариантах) меньше или 130. А в *списках 2* и *4* количество аффиксов с обязательным участием аффиксов сказуемости составляло 47. Здесь мы приводим только фрагменты *списков 1* и *2*, т. е. фрагменты *списков,* предназначенных для основ существительных с конечным мягким слогом (в Приложении эти же *списки 1* и 2 приведены полностью).

Известно, что конечный глухой согласный звук корня ***-қ, -к, -п*** переходит в звонкий, если присоединяемый к нему аффикс начинается с гласного, т.е. ***қ – ғ, к – г, п – б*** *(тарақ – тарағы, жүрек – жүрегі, доп – добы).* Поэтому в алгоритме задачи синтеза именной словоформы для автоматического словаря казахского языка при присоединении аффиксов к основе необходимо учесть воздействие последующего звука на предыдущий, т.е. учесть законы регрессивной ассимиляции. Здесь мы опускаем возможные варианты присоединения к форме единственного числа слов особых аффиксов, которые также зависят от последнего слога и звука слова в именительном падеже и осуществляются согласно закону сингармонизма.

Список 1 (фрагмент)

Список аффиксов, используемых для порождения словоформ от корневых и производных основ существительных с конечным мягким слогом (участие аффиксов сказуемости не обязательно)

Перечень аффиксов, соответственно типу слов-основ и конечных звуков (букв) у существительных в реестре АСКЯ					
№ п/п	Тип 1 на гласные, кроме ***у, и***	Тип 2 на буквы: ***у, и, й, р***	Тип 3 на буквы: ***м, н, ң***	Тип 4 на буквы: ***п, т, к, ш, с***	Тип 5 на буквы: ***ж, з, л***
	әке, келі, мәсі, мекеме	*үй, белбеу, темір, сәби*	*кілем, кескін, өлең*	*шөп, сүт, күрек, жүріс*	*тәж, жез, киіз, көз*
1	*ге*	*лерімнен*	*деріңіз*	*імнің*	*ің*
---	*де*	*лерімнің*	*деріңізбен*	*іміз*	*іңде*
20	*ден*	*леріміз*	*деріңізбенен*	*імізбен*	*іңді*
---	-------	*лерімізбен*	*деріңізге*	*імізбенен*	*іңдікі*
30	*лерден*	*лерімізбенен*	*деріңізде*	*імізге*	*іңдікімен*
---	-------	*лерімізге*	*деріңізден*	*імізде*	*іңдікіменен*
---	-------	*лерімізде*	*деріңізді*	*імізден*	*іңдікіне*
---	--------	*леріміздің*	*деріңіздің*	*іміздікімен*	*іңдікінен*
50	*лерім*	---------	*ім*	*іміздікіменен*	*іңдікінің*
---	--------	*лерімізді*	----------	----------	----------
---	--------	*леріміздің*	*імде*	*тікі*	*іңізбен*
100	*лерімен*	*леріңді*	----------	----------	----------
---	--------	---------	*іме*	*тікіне*	*іңізбенен*
120	*леріммен*	*леріңмен*	*мен*	*тікімен*	*іңізге*
---	--------	---------	----------	----------	----------
130	*сінің*	*іңіздікі*	*менен*	*тікіменен*	*іңізде*

Относительно составленных нами списков отметим, что мы далеки от мысли утверждать об окончательном их варианте как в количественном, так и в качественном отношениях. Также не исключаем, что синтез именных словоформ может быть осуществлен и другими путями. В любом случае предлагаемый нами вариант может послужить началом для осуществления компьютерного АСКЯ.

2.5. Об алгоритме программы – синтез именных словоформ казахского языка

Алгоритм программы синтеза именных словоформ казахского языка составлен с расчетом на четыре *списка*, о которых шла речь в предыдущем пункте. Для работы алгоритма обязательным условием является наличие *списка 1* и *2*, составленных для конечных мягких слогов именных основ, а также *списка 3 и 4*, рассчитанных для конечных твердых слогов слов из реестра АСКЯ. Мы уже отмечали, что разделение формообразующих списков на четыре разновидности задумано специально, чтобы довести до минимума процесс анализа структуры именной словоформы. Благодаря наличию указанных списков, работа алгоритма

Список 2 (фрагмент)

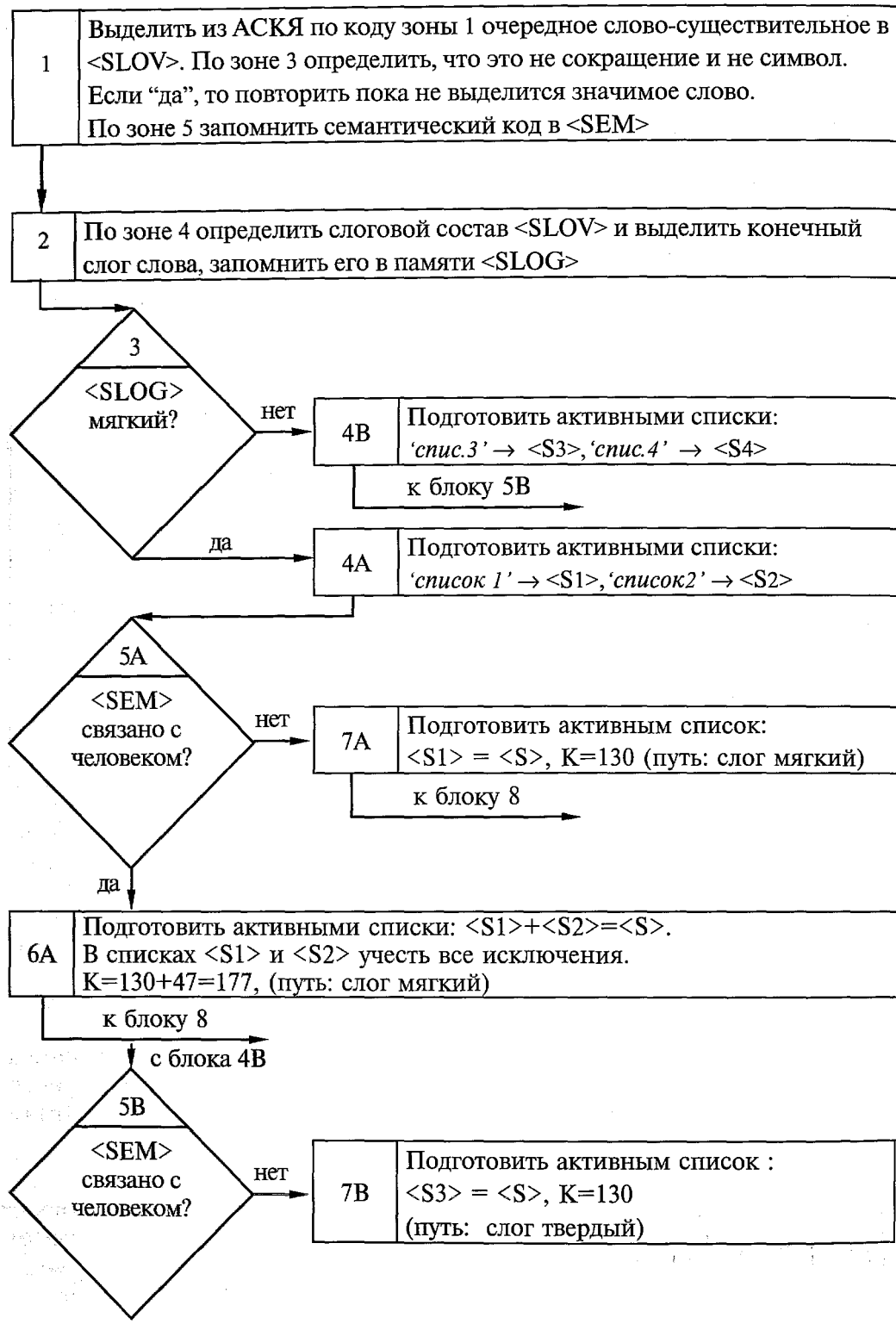
Список аффиксов, используемых для порождения словоформ от корневых и производных основ существительных с конечным мягким слогом (синтез именных словоформ с обязательным участием аффиксов сказуемости)

Перечень аффиксов, соответственно типу слов-основ и конечных звуков (букв) у существительных в реестре АСКЯ				
№ п/п	Тип 1 на гласные, кроме ***у, и***	Тип 2 на буквы: ***у, и, й, р***	Тип 3 на буквы: ***м, н, ң, ж, з, л***	Тип 4 на буквы: ***п, т, к, ш, с***
	кісі, кенже, кәсіпші, ене	*дәрігер, сәби, керей, кедей*	*көсем, келін, кербез, әйел*	*жігіт, білгіш, тентек*
1	*лерміз*	*імсіздер*	*дерміз*	*терміз*
2	*лермісіздер*	*імсің*	*дермісіздер*	*термісіздер*
3	*лерсіз*	*імсіңдер*	*дерсіз*	*терсіз*
4	*лерсіздер*	*імізбісіз*	*дерсіздер*	*терсіздер*
5	*лерсіңдер*	*імізсіздер*	*дерсіңдер*	*терсіңдер*
---	----------	----------	----------	----------
11	----------	*місіз*	*дерімсіңдер*	*терімсіңдер*
12	*леріміз*	*місіздер*	*деріміз*	*теріміз*
---	*лерімісіз*	*місің*	*дерімісіз*	*терімісіз*
15	*лерімісіздер*	----------	----------	----------
---	----------	*іңізбіз*	----------	*терісіз*
20	*лерімісіңдер*	*іңізбін*	*дерісіздер*	*терісіздер*
---	----------	*сіз*	*дерісіңдер*	*терісіңдер*
30	*лерісіңдер*	*сіздер*	*імбісің*	*імбісің*
---	----------	*імін*	*імсіз*	*імсіз*
46	*мбісің*	*імісіз*	*імсіздер*	*ісіздер*
47	*мсің*	*імісіздер*	*імсің*	*ісің*

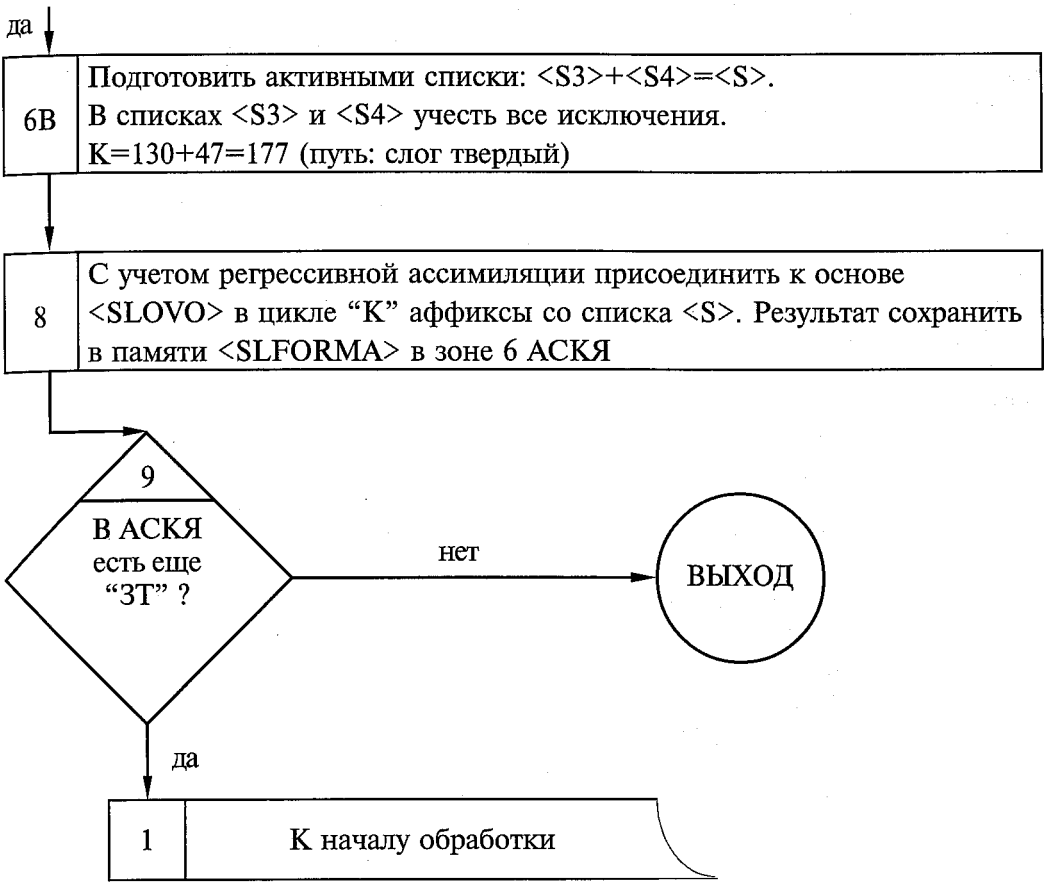
программы доведена до формальной стыковки выбранной именной основы из АСКЯ с аффиксами из соответствующих упомянутых списков. Некоторые проверки, имеющиеся в алгоритме, включены преднамерено из-за вынужденных исключений, которые появились при перегруппировке фонетических вариантов при простом склонении. Кроме того, при стыковке именной основы с очередным аффиксом из списка возникала необходимость учета законов сингармонизма и правил прогрессивной и регрессивной ассимиляции, которые также требуют небольшого анализа.

Ниже приводим в виде укрепненной *блок-схемы 1* алгоритм задачи синтеза именных словоформ казахского языка, которая даст общую картину поставленной перед нами задачи и послужит основанием для составления компьютерных программ для АСКЯ.

Блок-схема 1. **Алгоритм задачи синтеза именных словоформ**



Окончание блок-схемы 1



2.6. Синтез казахских глагольных словоформ для АСКЯ

Не все языковеды на должном уровне обращают внимание на количественную сторону образования словоформ от слов-основ определенных классов слов тюркских языков. В этом отношении следует отметить исследования Х. Ф. Исхаковой в области формальной морфологии тюркских языков. По ее данным, в татарском, турецком и узбекском языках от каждой основы-слова, относящегося к существительному, образуется 202 формы, от одной глагольной основы в татарском языке образуется 17 947 словоформ, в турецком – 11 390 и узбекском – 13 592 словоформ [136, с.7 и 17].

Статистическое исследование казахских глагольных словоформ на структурно-морфологическом уровне с использованием материалов частотных словарей по разным стилям казахского языка проведено К. Б. Бектаевым и С. Мырзабековым [36, 37]. По статистическим данным, их исследования, от каждой корневой или производной основы глагола аффиксальным путем может образоваться около 1150 глагольных форм, которые относятся к литературной норме казахского языка.

А если учесть формообразования глаголов с помощью союзов и других вариантов, то это число может достичь 1500 и более [37, с. 202-222]. Большая разница в числе словоформ от одной глагольной основы казахского языка по сравнению с такими фактами других тюркских языков объясняется спецификой казахского языка, где многие аффиксальные формы глагола считаются корневыми или производными основами глагола.

Авторы указанных статей обращают внимание и на статистику в отношении количества и порядка присоединения формообразующих аффиксов к глагольной основе. Например, одноаффиксная форма всретилась в 20 вариантах, двухаффиксная форма – 120, трехаффиксная – 310, пятиаффиксная – 372, шестиаффиксная – 78 и семиаффиксная форма зафиксировна в 24 вариантах [37, с.202].

Слово-основа казахского глагола в АСКЯ является вполне самостоятельным словом, представляющиv собой форму 2-го лица единственного числа будущего времени повелительного наклонения. В некоторых случаях к слове-основе присоединяется аффикс ***у*** (в виде *тұйық рай*). Глагольные основы в АСКЯ могут отличаться друг от друга и по морфологической структуре, и по количеству слогов. В модели АСКЯ будут фигурировать глагольные основы, главным формальным признаком которых является возможность прибавления отрицательной частицы: ***-ма, -ме, -ба, -бе, -па, -пе*** по законам сингармонизма.

Заметим, что возможность образования отрицательной формы глагола с помощью частиц ***-ма/-ме, -ба/-бе, -па/-пе*** является как бы особым признаком глагола, отличающим его о других частей речи казахского языка [163, с.288; 186, с.79]. В связи с этим, А. Хасенова отмечает, что указанные отрицательные частицы глагола, являясь формообразующими аффиксами, в то же время служат средством для определения границ глагольных основ от глагольных форм (грамматических значений) [141, с.55]. Например, для того чтобы узнать о том, приписанная к основе глагола морфема образует производную основу или только его форму с новым грамматическим значением, достаточно произнести это слово, добавив в конце соответствующую отрицательную частицу, выбрав ее согласно закону сингармонизма. Если при этом по значению не замечается нарушение отрицательной формы глагола, то проверяемое слово (часть слова до отрицательной частицы) считается глагольной основой (корневой или производной), а если происходит искажение отрицательной формы, то искомое слово есть глагольная словоформа.

Ввиду того, что большинство глагольных основ составляет производные основы, образованные путем аффиксации (синтаксический тип словообразования), то в предлагаемой модели АСКЯ они должны (во время поиска) формироваться для каждой основы глагола по правилам (алгоритму), которые должны быть указаны в зоне 6 словарной статьи.

Необходимо заметить, что аналитический тип словообразования, т.е. так называемые составные и сложные глаголы, состоящие из двух или более слов в модели АСКЯ, не будет рассматриваться (способ их формального формирования в АСКЯ требует отдельного исследования).

Отрицательные основы глагола не включены в АСКЯ, так как они так же, как и грамматические формы глагола, должны формироваться в зоне 6.

Различные глагольные словоформы образуют своеобразную систему, которую можно рассматривать как грамматические категории: 1) категория залога *(emic категориясы)*; 2) категория видов *(сипат категориясы)*; 3) категория положительности *(болымдылық)* и отрицательности *(болымсыздық);* 4) категория

наклонения *(рай категориясы);* 5) категория времени *(шақ категориясы);* 6) категория спряжения глагола по лицам и числам *(етістіктің жіктелуі)*; 7) категория причастия *(есімше);* 8) категория деепричастия *(көсемше);* 9) категория степени происхождения действия *(амалдың өту сипаты).*

Категория залога, выражая отношение действия к объекту и субъекту, считается лексико-семантическим признаком глагола, так как при этом пополняется и расширяется основное значение глагола. Поэтому считают, что при помощи залоговых аффиксов от глагольных корней или основ в казахском языке образуются производные глагольные основы. Ссылаясь на это, мы решили отнести их к словарной единице АСКЯ.

Производные глаголы, образованные от именных основ морфологичесим способом, т.е. при помощи аффиксов: ***-ла/-ле, -да/-де, -та/-те, -а, -е, -ай/-ей, -ар/-ер/-р, -ғар/-гер, -қар/-кер*** **и др. (*ой+ла, өткір+ле, түн+е, көб+ей, ақ+ар*** и т.д.), должны быть включены в АСКЯ так же, как и непроиводные глагольные основы.

Мы предлагаем формальный метод формообразования глаголов от глагольных основ (производные и непроизводные) посредством аффиксов, т.е. рассмотренные выше грамматические категории, такие, как видовые формы, наклонения, время, спряжения глагола, формы причастия и деепричастия (за исключением залоговых форм) синтезировать (порождать) программным путем, согласно грамматическим правилам современного казахского языка. Порождению производных форм алгоритмическим путем (компьютерным) благоприятствует тот факт, что в казахском языке формообразующие и словоизменяющиеся аффиксы присоединяются к слову в определенном порядке как, например, глаголы *көріспейсің* и *ойлатпайсыз*:

а) непроизводная основа *(көр)* + аффикс залога *(іс)* + аффикс отрицания *(пе)* + аффикс времени *(й)*+личный аффикс *(сің);*

б) именная основа *(ой)* + глаголообразующий аффикс *(ла)* + аффикс залога *(т)* + аффикс отрицания *(па)* + аффикс времени *(й)* + личный аффикс *(сыз).*

Руководствуясь этими положениями и разработанной методикой формообразования существительных, мы составили списки аффиксов, используемых для порождения словоформ от корневых и производных основ казахского глагола. При этом нами экспериментально выявлено 700 вариантов формоизменительных аффиксов как для глагольных основ с мягким конечным слогом в отдельности, так и для случая с твердым слогом. Для полноты охвата всевозможных формообразующих аффиксов казахского глагола мы хотим обратить внимание читателей на исследования К. Бектаева и С. Мырзабекова [36, 37, 200].

Так же, как и формообразующие списки для именных основ, экспериментально составленные нами формообразующие списки для глагольных основ должны постоянно подвергаться качественным и количественным изменениям в связи с нововведениями в структуре АСКЯ.

Приложении книги приведен список аффиксов синтеза словоформ от корневых и производных глагольных основ с конечным твердым слогом (*спис. 3*).

Глава третья

СЕМАНТИЧЕСКАЯ КЛАССИФИКАЦИЯ ЗНАМЕНАТЕЛЬНЫХ КЛАССОВ СЛОВ КАЗАХСКОГО ЯЗЫКА

3.1. Принципы семантической классификации лексики казахского языка

В соотношении "язык – реальная действительность" язык служит не только для обозначения предметов, фактов и явлений реальной действительности. Он, в то же время, является носителем обобщенного отражения действительности, органически соединяясь с соответствующими понятиями или отдельными их признаками, лежащими в основе семантических ценностей языковых единиц.

Общеизвестно, что лексические средства, наряду со звуковыми и грамматическими, составляют основу любого языка, в частности, и казахского. Слово, называя предметы, явления реального мира в их обобщенном виде, выступает в языке основным средством номинации. Являясь структурной ячейкой строя языка, оно имеет определенную морфологическую оформленность и грамматическую значимость в языке. Слово охватывает фактически весь объем языковых функций: номинативную (обозначения), сигнификативную (обобщения), коммуникативную (общения) и прагматическую (экспрессивного выражения и эмоционального воздействия). Словарные знаки слов таких классов, как существительные, глаголы и прилагательные, обладают двумя основными семиологическими функциями – наименования и описания. Совместимость их в пределах слова порождает назывные высказывания, предикативные и лексические синтагмы. Они являются тем основным звеном в словарном составе языка, той постоянной значимой основой, которая обеспечивает "переход" от языка к мысли, от системы к речи, от слов к синтагматическим единицам, большим, чем слова [289, с.82-100].

При семантической классификации лексики казахского языка (выделении в ее составе определенных лексико-семантических классов и подклассов) имеются трудности и специфика. Это одна из труднейших задач семантики, которая не имеет единого решения и в других языках мира. Количество таких классов и подклассов и их состав зависят от целей исследования, лингвистической школы и ряда других факторов.

Работ, посвященных исследованию казахского языка в области семантики, относительно мало. Здесь стоит отметить М. Оразова и его монографию "Қазақ тілінің семантикасы" и другие его работы [212–215], которые затрагивают многие вопросы семантики казахского языка. В теоретических и прикладных аспектах казахстанскими учеными проводятся различные исследования, затрагивающие семантические стороны

языковых единиц казахского и русского языков. Большую научную ценность имеют интерлингвальные исследования Э. Д. Сулейменовой по установлению семантических параметров слова в аспекте контрастивного анализа [275].

Особого внимания заслуживают научные изыскания З. К. Ахметжановой, посвященные функционально-семантическим полям русского и казахского языков. Применение функционального подхода, основных понятий теории функциональной грамматики к сопоставительному анализу типологически контрастных языков смело можно назвать началом нового этапа в развитии сопоставительного языкознания, так как здесь открывается возможность изучения системы языка в действии, в живом употреблении, т.е. сопоставительно-типологического исследования нескольких языков в функциональном аспекте. Кроме того, исследования З. К. Ахметжановой позволяют говорить о текстоорганизующей роли посессивности казахского языка и об отсутствии таковой функции у русского языка [18].

Можно было бы остановиться на исследованиях молодых ученых по отдельным проблемным вопросам семантики казахского языка [19, 95, 96, 156, 157, 168, 284, 310], но мы не ставим себе целью делать обзор работ, посвященных исследованиям семантики казахского языка. Эти труды необходимы нам для решения практических задач нашего исследования.

В самом общем виде можно выделить три принципа семантической классификации лексики казахского языка [58, 126, 215]: 1) денотативно-десигнативный; 2) парадигматический; 3) синтагматический.

Денотативно-десигнативный принцип учитывает онтологическое расчленение действительности. При этом слова объединяются в группы в соответствии с тем, что они обозначают в реальной действительности ("названия животных", "названия растений", наименования определенных классов предметов и т.п.). При *парадигматической классификации* слова объединяются в группы на основе тождества определенных семантических компонентов, выделенных в значении слов ("модальные глаголы", "каузативные глаголы" и т.п.). По синтагматическому принципу слова объединяются в лексико-семантические группы с учетом количества и качества валентностей классифицируемых слов [59; 215, с.189-192].

Исходя из целей нашего исследования, в основу семантической классификации положен *денотативно-десигнативный принцип*, хотя в некоторых случаях он будет дополняться двумя другими, перечисленными выше.

3.2. Семантическая классификация имен существительных *(зат есім)*

Как известно, основным критерием разграничения имен существительных казахского языка на отдельные группы, как и в других языках, является, прежде всего, анализ их прямого номинативного значения, детерминированного типом денотата и сигнификата. При этом, "слово" можно рассматривать двояко, т.е. как единицу первичного означивания в системе номинативных средств языка и как единицу, семантически реализованную в речи. Нами различаются две модификации "денотатов": денотат как объект обозначения единичного предмета или класса предметов, а также денотат как предмет референции знака в речи.

А. А. Уфимцева, в книге "Типы словесных знаков" [289, с.102], трактует сигнификат в шести понятиях как компонент знакового значения слов, что, как нам кажется, справедливо и для казахского языка:

1. Содержание полного конкретного понятия, противопоставленное его объему (денотату), как это имеет место у имен классов конкретных или единичных предметов, явлений. Например: *қала (город), жел (ветер), кітап (книга)* и т.п.

2. Абстрактное понятие признака, свойства, действия, отношения и т.п., в высшей степени обобщенное и непротивопоставленное предметному представлению. Например, *данышпандық (гениальность), ер жету (зрелость), надандық (тупость), қараңғылық (темнота), суықтық (холодность), жан тыныштық (спокойствие)* и т.п.

3. Абсолютные, в высшей степени обобщенные понятия, выражающие мыслительные категории. Например, *біртуар (феномен), белгі (знак), нышан (признак), қасиет (свойство), сан (количество).*

4. Абстрактное понятие, обозначающее категории реального мира и выступающее родовым по отношению к другим. Например, *сезім (чувство), оқиға (событие), қозғалыс (движение), әрекет (действие), сұйықтық (жидкость)* и т.п.

5. Единичное понятие, представляющее собой "мыслительный конструкт" ирреального предмета. Например, *кентавр, леший, русалка* и т.п. В казахском языке такие, как *диуана, пері, жын-шайтан* и т.п.

6. Метаязыковые понятия, термины различных областей знаний (науки, техники, искусства и т.п.).

Из перечисленного вытекает, что имена с денотативным и денотативно-сигнификативным значением относятся к разряду имен, называемых конкретной лексикой, а слова с сигнификативной и сигнификативно-денотативной основой значения составляют абстрактную лексику любого языка. Необходимо заметить, что определяющим для семантики именных лексем является обязательное наличие в их знаковом значении денотата и сигнификата. Если словесный знак служит специализированным названием (термином) в какой-либо узкой предметной области, то доминирует денотат над сигнификатом. В случае, когда словесный знак служит названием абстрактного понятия, то, наоборот, или сигнификат доминирует над денотатом, или словесный знак имеет чисто сигнификативное значение [289, с.103].

Именами существительными в казахском языке, так же, как и во многих тюрских языках, являются различные наименования предметов и понятий в широком смысле слова. Это, например, различные имена собственные, такие, как имена людей, наименования городов, рек, озер и т. д.; нарицательные имена – названия человека: *кісі (человек), шеше (мать), әйел (женщина);* названия животных; растений; названия различных веществ и предметов быта; научные, технические и другие термины; наименования разных обобщенных отвлеченных понятий: *ел (народ), ру (род), тағдыр (судьба)* и т.д.; наименования действий, явлений, состояний: *жүріс (ходьба), хабар (известие), ұйқы (сон)* и т.д.; названия результатов действия: *жиналыс (собрание; ср. жинау – собираться), кеспе (лапша; ср. кесу – резать);* названия орудий (инструментов) действия: *бұрғы (сверло; ср. бұрау – винтить, завинчивать), сыпырғыш (метла, веник), ара (пила)* и т.п.

В отличие от европейских языков, в казахском языке имена существительные не различаются грамматически по признаку одушевленности или неодушевленности. В семантическом отношении только наименования людей различаются как

одушевленные, отвечая на вопрос *кім? (кто?),* а все остальные (живые и не живые) отвечают на вопрос *не? (что?).* В случае необходимости различения лиц мужского и женского пола животного мира используются понятиями пола по значению самого слова [163, 164, 276, 327].

Корневые имена существительные в казахском языке в основном совпадают с корнем и, обладая полнотой значения, соответствуют самостоятельному слову. Последнее намного облегчает процесс автоматического расспознавания их в тексте. Термин "производная основа" может быть применен в казахском языке, главным образом, в отношении глаголов и отглагольно-именных образований. Таким образом, имена в казахском языке являются самостоятельной лексико-грамматической категорией и в предложении могут употребляться без дополнительных аффиксов [327].

Как пишет М. Оразов, среди слов казахских существительных самыми частыми по сложности их значений являются нарицательные имена. Эти имена существительные по отношению к значению имеют в языке изменчивый характер, т.е. иногда у них превалирует денотативное значение, а иногда сигнификативное или приобретается как бы суммарное значение – и денотативное, и сигнификативное. Примером может служить использование слов, которые применяются в метафорическом и метомических значениях. Безусловно, когда слова употребляются в переносном значении, то изменяется его и денотативное, и сигнификативное значение. Для убедительности изложенного можно сравнить слово "нар" в двух следующих предложениях из стихов: *қырық балам бар еді, қырқы да бірдей нар еді* (Батырлар жыры). *Айырдан туған жампоз бар - нарға жүгін салғызбас* (Махамбет) [215, с.98]. В первом предложении слово "нар" применено в переносном значении, т.е. как сильный, "боевой", а во втором – в прямом значении, как одногорбный верблюд. Их денотативные и сигнификативные значения несравнимы.

В количественном отношении тематические группы имен существительных в любом языке, в том числе, и в казахском, довольно велики. М. Оразов в книге "Қазақ тілінің семантикасы" приводит в виде примера 8 разновидностей таких тематических групп [215, с.191]: 1) родственные термины *(туысқандық атаулар);* 2) человек и названия его части тела *(дене мүше атаулары);* 3) названия домашных животных *(төрт түлік мал атаулары);* 4) названия одежды, обуви человека *(киім-кешек атаулары);* 5) название птиц *(құс атаулары);* 6) название диких животных (жабайы аң атаулары); 7) название обиходных предметов *(ыдыс-аяқ атаулары);* 8) названия болезни человека *(адам баласына тән ауру атаулары).*

С учетом целей нашего исследования, анализируемого материала, а также отдельных идей, изложенных в работах [11, 18, 19, 142, 215 и др.], нами выделены следующие семантические классы существительных, которые приведены в *табл. 3.* В нашем исследовании в основу семантической классификации казахских существительных положено деление реальной действительности на определенные зоны, сегменты [126, 142, 215]. Как видно из этой *табл.*, денотативно-десигнативный принцип использован здесь при выделении наиболее крупных семантических классов существительных. Более мелкие лексико-семантические группы, например "части тела человека", "воинские звания", "дни недели" и др., связаны парадигматическими отношениями.

Таблица 3. **Семантические классы существительных, использованных в модели порождения казахского текста**

Сегменты реальной действительности	Наименование семантического класса	Код класса
1	2	3
1. Наука	Общефилософские понятия	ЗТ101
	Общенаучные понятия	ЗТ102
	Общемедицинские понятия	ЗТ103
	Основные понятия психологии	ЗТ104
	Основные понятия нейропсихологии	ЗТ105
	Основные понятия логики	ЗТ106
	Основные понятия геометрии	ЗТ107
	Основные географические понятия	ЗТ108
	Основные понятия языкознания	ЗТ109
	Основные понятия литературоведения	ЗТ110
	Основные понятия изобразительного искусства	ЗТ111
	Основные понятия музыкального искусства	ЗТ112
	Основные юридические и уголовные понятия	ЗТ113
	Основные понятия, связанные с театром	ЗТ114
	Понятия, связанные с авиацией	ЗТ115
	Понятия, связанные со строительством	ЗТ116
	Понятия, связанные с устройством отдельного механического объекта	ЗТ117
	Информационные понятия	ЗТ118
	Собирательные понятия	ЗТ119
	Основные понятия электроники и электротехники	ЗТ120
	Административные понятия	ЗТ121
	Имена организаций, газет, журналов и т.д.	ЗТ122
	Абстрактные имена – мыслимые образы	ЗТ123
	Оценка	ЗТ124
2. Природа	Предметы природы	ЗТ201
	Явления природы	ЗТ202
	Предметы природы – артефакты	ЗТ203
	Состояния природы	ЗТ204
	Понятия, описывающие структуру предметов и части предметов	ЗТ205
	Вещества, металлы	ЗТ206
	Действия предметов природы	ЗТ207

Продолжение табл. 3

1	2	3
3. Человек и его мир	Человек и его части тела	ЗТ301
	Части лица и головы человека	ЗТ302
	Состояния и ощущения, специфичные для человека	ЗТ303
	Действия, спецефичные для человека	ЗТ304
	Понятия, используемые при характеристике человека	ЗТ305
	Предметы одежды и обуви человека и понятия, с этим связанные	ЗТ306
	Предметы-артефакты, используемые человеком в целях гигиены, для украшения и т.п.	ЗТ307
	Объекты, используемые для передвижения человека и понятия, с этим связанные	ЗТ308
	Предметы питания человека	ЗТ309
	Понятия, связанные с жилищем человека	ЗТ310
	Понятия, связанные со смертью и захоронением	ЗТ311
	Имена, отчества и фамилии людей	ЗТ312
	Национальность человека	ЗТ313
	Предметы-сосуды, которыми пользуется человек, и связанные с этим понятия	ЗТ314
	Предметы, используемые на отдыхе, охоте, рыбалке, и понятия, с этим связанные	ЗТ315
	Наименование гражданских специальностей, должностей, степеней, званий	ЗТ316
	Предметы-артефакты, используемые при учении, труде и т.п.	ЗТ317
	Понятия, связанные с взаимоотношением людей	ЗТ318
	Понятия, связанные с родственными отношениями между людьми	ЗТ319
4. Животный мир	Награды человека	ЗТ320
	Крупные домашние животные и части их тела	ЗТ401
	Мелкие домашние животные и части их тела	ЗТ402
	Крупные дикие животные и части их тела	ЗТ403
	Насекомые и части их тела	ЗТ404
	Птицы и части их тела	ЗТ405
	Действия, спецефичные для животных	ЗТ406
	Предметы питания животных, птиц и насекомых	ЗТ407
5.Растения	Растения и их части	ЗТ501
	Деревья и их части	ЗТ502
6. Время	Временные понятия	ЗТ601
	Дни недели	ЗТ602

Окончание табл. 3

1	2	3
7. Религия	Понятия, связанные с религией	ЗТ701
8. Война и понятия, с ней связанные	Предметы боевой техники, оружия и их части	ЗТ801
	Наименование военных специональностей	ЗТ802
	Воинские звания	ЗТ803
	Понятия, связанные с оказанием помощи во время войны	ЗТ804
	Названия воинских подразделений	ЗТ805
	Понятия, связанные с войной	ЗТ806
9. Учеба и труд	Организация учебы, труда	ЗТ901
	Понятия, связанные с обучением и трудом	ЗТ902
10. Меры	Меры длины, веса, скорости и т.п.	ЗТ911
11. География	Имена населенных пунктов	ЗТ921
	Имена стран, материка, частей света, островов, гор и др.	ЗТ922
	Имена рек, морей, озер, проливов	ЗТ923

В последней графе *табл. 3* приведены принятые в модели порождения коды семантических подклассов существительных. ЗТ – сокращение от *зат есім* (существительное).

3.3. Семантическая классификация глаголов казахского языка *(етістік)*

3.3.1. Знаковое значение глагольных лексем

Если имена существительные отражают и обозначают предметы или явления, то глаголы – отношения между словами. Поэтому значения существительных более автономны и абсолютны, а значения глагольных лексем относительны и детерминируются многими факторами. В глагольных лексемах, выражающих своей семантикой понятие отношения, их взаимодействие как бы выносится за рамки глагольной лексемы и переносится в область синтагматических отношений действия к его объекту или субъекту (или одновременно). Поэтому соотнесенность глагольных лексем с предметным рядом предполагает наличие смысловых связей глагола с предметными именами.

Знаковое значение глагольных лексем представляет собой сложный феномен, детерминируемый тремя факторами: 1) соотнесенностью с реальной действительностью, которая представляет собой в данном случае не мир предметов, а мир их отношений, действий и состояний; 2) категориальной семантикой сочетающихся с глаголом предметных имен; 3) типом смысловых отношений между действием, его субъектом и объектом, а соответственно типом моделей "объект – действие", "действие – объект", "субъект – действие – объект" [289, с.113].

По сравнению с именами существительными класс глагольных лексем представляет большие трудности при попытках их классификации по характеру их знакового значения. Во-первых, собственно лексическое (индивидуальное)

значение глагольных слов осложенено рядом семантических признаков, выражающих способ протекания действия, обстоятельства его свершения, спецификой видовременных характеристик глагольного действия, наличия модальных и экспрессивных наслоении. Во-вторых, большое затруднение в определении типа глагольной семантики заключается в многоаспектности каждого семантического признака: присущ он только данной лексеме или целой группировке или даже целой категории глагольных лексем.

Традиционно разграничение глаголов по их семантике на три следующих класса: 1) глаголы действия; 2) глаголы состояния; 3) глаголы процессов [289, с.120]. У каждой лексико-семантической группы имеются парадигматческие параметры, семантические признаки. Так, для глаголов "движения", довольно хорошо изученных во многих языках, выделяются следующие семантические признаки: "направленные движения", "скорость движения", "манера передвижения", "средства" или "среда передвижения" и т.п. Для семантики эмотивных глаголов существует иная шкала. Среди них выделяют глаголы воздействия, глаголы восприятия, глаголы оценки или субъективного отношения и т.п. [289, с.148].

М. Оразов пишет, что не конкретность лексического значения глагола в основном зависит от его денотативного и сигнификативного значения [215, с.99]. Когда речь идет о денотативном значении глагола, он имеет в виду движение предмета, а также физическое действие *(амал-әрекет)* и состояние *(қалпы)*. Если учесть, что движение происходит в какое-то определенное время и каком-то месте, то здесь имеется в виду присутствие субъекта, который осуществляет физическое действие, а также известна цель передвижения и при этом лексическое значение глагола дополняет множество других элементов. В качестве примера автором книги приведено следующее предложение: *Шаңғышылар селоға кіргенде көз байланып та қалған еді (М. Сүндетов) /Когда лыжники вошли в село, уже наступило сумерки/.* В этом предложении рассматривается структура значения глагола *кір (входи)*: 1) наличие движения, 2) движется некоторый предмет или человек (субъект), 3) движение происходит на земле, 4) переход через некоторую границу (вошли на территорию села), 5) имеется элемент значения войти "во внутрь". Эти значения определяются благодаря отношениям и связям слов в данном предложении. Таким образом, действие *кіру (входить)* происходит на лыжне (на снегу) движением лыжников, *кіру* означает вхождение на территорию, где проживают люди (в село). Учитывая сочетание слова *кір* с другими словами *(село, шаңғышылар),* мы узнаем направление движения, что оно происходит *"извне в село"*. Данный же глагол *кір*, сочетаясь с другими словами, может принять абсолютно другое значение. Например: *Мен 1962 жылы партия ұйымына кірдім.* В этом предложении слово *кір* уже не означает движение, а понятие о том, что он стал членом партийной организации. Последний пример дает основание полагать, что значение слова непосредственно не зависит от значения словосочетаний (от контекста в широком смысле) в предложении. В первом примере глагол *кір* сочетается с объектом, означающим местность, поэтому его граница является основой понимания лексического значения из сем движения. Во втором примере глагол *кір* сочетается с понятием, т.е. названием организации, здесь сема глагола *кір* единственная – стать членом данной организации. Изменение значения некоторых глаголов требует очень тонкого подхода при анализе их семантики. Например, глаголы движения в одних сочетаниях входят в структуру глагола общения, а в других – в структуру глагола мышления. Сравните: *үйге келу (придти домой), қарсы келу* (возразить), *ойға келу* (вспомнить),

тілге келу (разговаривать) и др. В структуре предложения глаголы образуют отношения с определенными словами. Такие отношения происходят между субъектом или объектом и движением, между физическим действием *(амал-әрекет)* и объектом или субъектом. Когда в предложении глагол "требует" связи с другим словом, то он с этим словом имеет связь в отношении и значения, и формы (валентность глагола).

3.3.2. Тематические и лексико-семантические классы казахских глаголов

Основанием для отнесения слов к различным тематическим и лексико-семантическим классам являются обобщающие *(жалпылаушы)* семы. Необходимо отметить, что в индо-европейских языках имеется определенный опыт в разделении на лексико-семантические классы отдельных частей речи, но в отношении тюркских языков, в частности казахского, подобные исследования начали проводиться только в последнее десятилетие (большинство из них имеет сопоставительный с русским языком характер).

В настоящее время защищен ряд кандидатских диссертаций по описанию следующих лексико-семантических групп глаголов казахского языка: особенности глаголов физиологического воздействия [168]; семантико-грамматические особенности глаголов речи [156]; семантика глаголов ориентированного положения в пространстве [95]; семантика глаголов движения [310]; эмотивные глаголы [130]; глаголы физического восприятия [288]; глаголы созидания [284] и др.

Интересные данные в этом плане получены М. Оразовым в уже упомянутых его работах. Нами использована его семантическая классификация отдельных частей речи для решения конкретной задачи – разработки процедуры порождения казахского текста.

Рассматривая проблему семантической классификации глагольной лексики, нельзя оставить без внимания такую категорию, отражающую одну из форм всеобщей связи явлений и фактов объективной действительности, как категория каузальности (причинности).

В исследованиях казахского языка категория каузальности в ее полном объеме не была объектом специального изучения. Каузальные отношения в языке рассматривались в пределах отдельно взятых лингвистических задач. Например, на морфемном уровне: формы повелительного наклонения глагола (Н. Сауранбаев, А. Исхаков, А. Калыбаева-Хасенова, С. Исаев, Н. Оралбаева, Е. Кажибеков, И. Маманов, К. Неталиева, К. Ищанов, О. Жармакин, Л. С. Дуйсенбекова и др.), предложные формы глагола (В. А. Исенгалиева, Г. А. Мейрамов, Р. Г. Баяминова и др.). На лексическом уровне явление каузации позволяет выделить семантические типы каузативных глаголов (М. Оразов, К. К. Дуйсекова, Т. Ж. Токтарова и др.). На синтаксическом уровне изучались сложноподчиненные предложения и бессоюзные сложные предложения с причинно-следственным характером отношений между составляющим их частями (М. Б. Балакаев, Т. Кордабаев, Н. Х. Демесинова, К. Есенов и др.).

В последнее годы научная школа, возглавлямая профессором З.К.Ахметжановой, намного продвинула вперед функционально-семантическое направление исследования в сопоставительном плане изучения казахского и русского

языков [9, 18-20, 169, 285]. При семантической классификации каузативных глаголов казахского языка мы пользовались данными работы Г. М. Алиной "Каузальные отношения в системе русского и казахского языков" (канд. дис., и др.) [9].

Необходимо отметить, что недостаточно изучена и еще одна группа глаголов казахского языка – бытийные глаголы. В существующих классификациях лексико-семантических групп казахского языка бытийные глаголы не выделяются как отдельная группа [169, с.46-52]. М.С.Койшибаева в статье "К вопросу о бытийных глаголах в русском и казахском языках" останавливается на точке зрения казахстанских ученых по данному вопросу. Автор статьи отмечает, что в казахском языке глаголы с чисто бытийным значением, такие, как *еді (екен, емес), бол,* характеризуются, в основном, как вспомогательные глаголы, которые "самостоятельно не употребляются, но образуют описательные формы времен и наклонений" основных глаголов в предложениях [185]. Согласно мнению А. Ыскакова, такие глаголы, образованные от корня ***-е** (еді, екен, емес)*, представляют собой неполные слова, которые утратили лексическое значение либо полностью, либо наполовину и играют в предложении вспомогательную роль, придавая грамматическое значение отдельным словам (*бала еді, қырық-отыздар шамасында екен, қара-торы ала емес*) [327, с.248].

Как отмечает В. А. Исенгалиева, "глагол *быть (бол)* двуфункционален. Он выступает в роли вспомогательного вербализующего компонента в составном глаголе, придавая ему процессуальное значение, и, в качестве самостоятельного глагола, имеющего комплекс своих значений, например *кеше жиналыс болды*" [135, с.55].

Относительно модально-предикативных слов *бар/жоқ* Н. Джанашина пишет: "Не вполне адекватно, но наиболее точно значение *var/yjk* можно определить как существование/несуществование" [97, с. 60-77].

Учитывая изложенные выше подходы к классификации глагольной лексики и ориентируясь на практическое их использования в нашей модели порождения казахского текста, мы предлагаем свою версию семантических классов казахских глаголов *(табл. 4).*

Таблица 4. **Семантические классы казахских глаголов, используемых в модели порождения текста**

Семантический подкласс глагола	Примеры	Код
1	2	3
Семантический класс: 1. Глаголы существования		
1.Бытийные глаголы: *Бар*	*Менде жақсы оқуға мүмкіншілік* ***бар****, бірақ оған тұқ ынтам* ***жоқ.***	ET101
Жоқ	*Оның жасы әлі көмелетке толған* ***жоқ.***	ET102
Бол (болу, болмау)	*Сөзіңді айтып* ***бол.*** *Кеше жиналыс* ***болды***	ET103
Е (еді, екен, емес)	*Алматы Қазақстанның астанасы* ***еді.*** *Өзі бір қызық адам* ***екен.*** *Оның аты қара-торы* ***емес.***	ET104

Продолжение табл. 4

1	2	3
Семантический класс: 2. Глаголы физических действий		
1. Глаголы физического изменения: созидания и производства, соединения и деления и т.д.	*Жасау, соғу, тігу, тұрғызу, салу, бөлшектеу, екшеу, ыдырату* и т.д.	ET201
2. Глаголы активных физических действий: разрушения и порчи	*Қирату, бүлдіру, бұзу, сындыру, құлату* и т.д.	ET202
3. Глаголы эмоциональных физических действий	*Аузы ашылу, мәз болу, күлу, ыржиу, шапалақтау, секіру (қуанғаннан)*	ET203
4. Глаголы употребляющиеся при приеме пищи, свойственных для человека и животного мира	*Ішу, жеу, ұрттау, жұту, асау, дәм тату, сораптау* и т.д.	ET204
5. Глаголы прочих физических действий, свойственных человеку и животному миру	*Жүгіру, секіру, ырғу, құсу, таптау, жылжу, қыймылдау* и т.д.	ET205
6. Глаголы физических действий, описывающих неживую природу	*Күркіреу, сарқырау, арсылдау, сілкіну* и т.д.	ET206
7. Глаголы прочих физических действий, свойственных неживой природе	*Ағу, бату (күн), құлау, қайнау, діріддеу, ысу, суу, гүрсілдеу* и т.д.	ET207
8. Глаголы, связанные со звуками окружающей среды, глаголы звуковыражения	*Тарсылдау, сатырлау, сарқырау, ызыңдау, гүрсілдеу, қоңқылдау* и т.д.	ET208
9. Глаголы физических действий, свойственных только животному миру	*Ұлу, ұру, ырылдау, мөңку, мөңіреу, жорту, шабу* и т.д.	ET209
Семантический класс: 3. Глаголы движения		
1. Глаголы, означающие движения, происходящие на земле	*Жүру, жүгіру, еңбектеу, домалау,сырғанау* и т.д.	ET301
Семантический класс: 3. Глаголы движения *(продолжение)*		
2. Глаголы движения, происходящие на воде (в воде)	*Жүзу, сүңгу, малту, қалқу* и т.д.	ET302
3 Глаголы движения, происходящие в воздухе	*Ұшу, самғау, қалықтау, шүйілігу, зымырау* и т.д.	ET303
4. Глаголы движения, означающие направленное движение	*Өрлеу, ылдилау, келу, кету, жақындау, ұзау*	ET304

Продолжение табл. 4

1	2	3
5. Глаголы движения, означающие хаотичность движения	*Аралау, шарлау, кезу, қыдыру, атқу* и т.д.	ET305
Семантический класс: 4. Глаголы, означающие отношения между людьми, животными и неживыми		
1. Глаголы выравнивания	*Алу, қосу, теңестіру, салыстыру, бөлу* и т.д.	ET401
2. Глаголы принадлежности	*Иемдену, меншіктеу, иелену, баурау* и т.д.	ET402
3. Глаголы приветствия и прощания	*Амандасу, сәлемдесу, көрісу, қоштасу* и т.д.	ET403
4. Глаголы почтения	*Құрметтеу, қадірлеу, іздеттеу, сыйлау* и т.д.	ET404
5. Глаголы властвования	*Басқару, әмір ету, бопсалау, өкім ету* и т.д.	ET405
6. Глаголы подчинения	*Көну, бас иіу, ырқына көну, бағыну* и т.д.	ET406
Семантический класс: 5. Глаголы психической деятельности		
1. Глаголы ощущения: а) глаголы ощущения в общем смысле	*Сезу, сезіну, байқау, абайлау, аңғару, пайымдау, түйсіну* и т.д.	ET501
б) глаголы видения	*Көру, көз салу, көз тастау, үңілу, бажырайу* и т.д.	ET502
в) глаголы вкусовые	*Кермектену, ашқылтымдану, тату* и т.д.	ET503
г) глаголы осязания (связанные с запахом)	*Бықсу, мұңку, аңқу, сасу, борсу, қаңсу* и т.д.	ET504
д) глаголы осязания, связанные с порами тела человека и животных	*Сыздау, ашу, қышу, тызылдау, сырқырау* и т.д.	ET505
е) глаголы, связанные со слухом	*Тыңдау, есіту, құлақ салу, дауыстау, естілу* и т.д.	ET506
2. Глаголы мышления: а) глаголы мышления, связанные с процессом мышления вообще	*Ойлау, пайымдау, ой жүгірту, ақылға салу, болжау, шамалау* и т.д.	ET507
б) глаголы предположения (приблизительности)	*Шамалау, болжау, тұспалдау, межелеу* и т.д.	ET508
в) глаголы, связанные с выводами, заключениями, обобщениями	*Қорытындылау, шешілу, кесіп айту* и т.д.	ET509
г) глаголы доверия	*Сену, достасу, дос көру, жақындасу, сенім арту.*	ET510
д) глаголы запоминания	*Ескер, ұмытпа, есте сақта, зерделе қалдыру ...*	ET511

Продолжение табл. 4

1	2	3
3. Глаголы психического воздействия	*Қуанту, күлдіру, қорқыту, ойландыру* и т.д.	ET512
4. Глаголы внимания и желания	*(жақсы, ақылды, сыпайы, ұстамды)болу* и т.д.	ET513
5. Глаголы активной психической деятельности	*Айқайлау, ұрысу, зар жылау, қарқылдау* и т.д.	ET514
6. Глаголы эмоциональные:		
а) глаголы восхищения	*Таңдану, қуану, таңырқау, аузы ашылу* и т.д.	ET515
б) глаголы радости	*Шаттану, қуану, мәз болу ...*	ET516
в) глаголы удовлетворения	*Рахаттану, қанағаттану, шаттану* и т.д.	ET517
г) глаголы эмоционального переживания (скорби)	*Қапалану, қайғыру, қабырғасы қайысу* и т.д.	ET518
7. Глаголы говорения: а) информативные	*Айту, хабарлау, сөйлеу, баяндау* и т.д.	ET519
б) связанные с процессом собеседования	*Сөйлесу, әңгімелесу, сырласу, сұбхаттасу* и т.д.	ET520
в) употребляющиеся в процессе “вопрос-ответ”	*Сұрау, жауап беру, кінәласу, беттесу* и т.д.	ET521
г) связанные с эмоциональным состоянием человека	*Мыңқылдау, көку, оттау, шалқу, лепіру* и т.д.	ET522
Семантический класс: 6. Глаголы физического состояния и нормы		
1. Глаголы, специфичных для элементов неживой природы (изменение цвета)	*Жылтырау, жарқырау, ағару, қарайу, сарғайу, көгеру, сәулелену* и т.д.	ET601
2. Глаголы физического сохранения	*Сақталу, бұзылмау, ашымау, қартаймау...*	ET602
а) означающие внутренние психологические состояния человека	*Қасарысу, ашулану, қайғыру, толқу, қуану, ойлану, армандау* и т.д.	ET603
б) означающие субъективные черты человека (наружные)	*Кісімсу, адамсу, баласыну, сасқалақтау* и т.д.	ET604
3. Глаголы, передающие изменение объема предмета, вещества	*Шөгу, ұзару, солу, толысу, зораю, кеңею* и т.д	ET605
4. Глаголы, передающие изменение предмета в количественном отношении	*Азаю, көбею, молаю, ортаю, қосылу* и т.д.	ET606
5. Глаголы, означающие изменение внутренних свойств человека или предмета	*Іру, шіру, үгілу, тозу, босау, өзгеру* и т.д.	ET607

Продолжение табл. 4

1	2	3
6. Глаголы, употребляющиеся в связи с привычками людей	*Әдеттену, үйрену, дағдылану, жалқаулану ...*	ЕТ608
7. Глаголы, связанные с процессом рождения и роста, свойственных человеку или животным	*Туу, өсу, өну, ержету, жігіт болу* и т.д.	ЕТ609
Семантический класс: 7. Глаголы психического состояния		
1. Глаголы, свойственные только человеку	*Қайғыру, күңірену, ренжу, күлу, жылау, қуану*	ЕТ701
2. Глаголы, связанные с природными явлениями	*Азынау, гүілдеу, борандату, желдету* и т.д.	ЕТ702
3. Глаголы, свойственные человеку и особям животного мира	*Қайғыру, қуану, ашулану, ызалану* и т.д.	ЕТ703
4. Глаголы, приписываемые элементам неживой природы	*Қайғыру, мұңаю, күңірену, қуану, жадырау...*	ЕТ704
5. Глаголы, свойственные человеку, особям животного мира и риписываемые элементам неживой природы	*Қайғыру, мұңаю, жадырау, күңірену, қуану, күлімсіреу* и т.д.	ЕТ705
Семантический класс: 8. Фазисные и модальные глаголы		
1. Начало действия	*Бастау, кірісу, жасағалы-, кіріскелі-* и т.д.	ЕТ801
2. Продолжения действия	*Жалғастыру, созу, атқару, -ып, -іп (отыр)...*	ЕТ802
3. Конец действия	*Аяқтау, қою, бітіру, тоқтату, болу* и т.д.	ЕТ803
Модальные глаголы состояния: *Отыр*	*Балалар сабақ оқып отыр.*	ЕТ804
Жатыр	*Оқушылар мектептен келе жатыр.*	ЕТ805
Тұр	*Әнші өлең айтып тұр.*	ЕТ806
Жүр	*Оқымыстылар космосты зерттеп жүр.*	ЕТ807
Семантический класс: 9. Каузативные глаголы		
1. Собственно-каузативные глаголы: - понуждение	*Мәжбүрле-, күште-, көндір-, душар ет-* и т.д.	ЕТ901
- вызывание, причинение	*Келтір-, туғыз-, оят-, тигіз-, қоздыр-* и т.д.	ЕТ902
- побуждение	*Түрткі бол-, көндір-, көңілін аудар-, тарт- ...*	ЕТ903
- обусловливание	*Жеткіз-, әкел-, апар-, түсір-* и т.д.	ЕТ904

Продолжение табл. 4

1	2	3
- воздействие	*Ет-, уста-, тұтын* и др.	ЕТ905
- допушение	*(мүмкіндік) бер-* и др.	ЕТ906
2. Частно-каузативные глаголы Глаголы созидания объектов: в содержании способ действия: - соединяя, скрепляя части	*Тік-, астарла-, тепші-, жапсыр-, құра-* и т.д.	ЕТ907
- путем одновременной деформации материала и его уплотнения, соединения	*Өр-, иір-, тоқы-, ес-, иле-, пысықта-, қарпы-, шират-* и т.д.	ЕТ908
- обрабатывание определенным образом поверхность чего-либо	*Еге-, жаңқала-, жаны-, сүргіле-, қайра-* и т.д.	ЕТ909
- инструмент действия	*Бұрғыла-, қаз-* и др.	ЕТ910
- свойство материала	*Құй-, жапсыр-, үрле-* ...	ЕТ911
- строить, воздвигать	*Сал-, тік-, қала* ...	ЕТ912
- приготовление пищи	*Пісір-, қуыр* и т.д.	ЕТ913
Глаголы уничтожения объекта в содержании способ действия:		
- огнем	*Өрте-, жандыр-* и т.д.	ЕТ914
- атакуя с воздуха	*Бомбыла-, атқыла-* ...	ЕТ915
- взрывом	*Қопар, құлат* ...	ЕТ916
- ломая	*Бұз-, қырат-, сындыр-*...	ЕТ917
- стремительно, с силой	*Шақ-, талқан қыл* ...	ЕТ918
- с яростью, с силой, до основания	*Астан-кестен қыл-, тас-талқан ет* и т.д.	ЕТ919
3. Глаголы со значением рождения	*Босандыр-, тап-, шағы-лыстыр-, төлдет-* и т.д.	ЕТ920
4. Глаголы со значением лишить жизни кого-либо	*Өлтір-, мерт қыл-, бауызда-, ас-, ат-* и т.д.	ЕТ921
5. Глаголы со значением изменения объектов в содержании способ действия: а) изменять первоначальную форму	*Бүкте-, қайыр-, бүк-, жалпайт-, ез-, мыжғыла-, умажда-* и др.	ЕТ922
б) придавать определенную форму	*Пішінде-, дөңгелекте-, үй-, түй-, өр-* и др.	ЕТ923
6. Глаголы, отношения объекта к признакам существования *Количественная характеристикака:* а) увеличения – уменьшения	*Көбейт-, молайт-, ұлғайт-, азайт-* и т.д.	ЕТ924

Продолжение табл. 4

1	2	3
б)удлинения – укорачивания	*Ұзарт-, қысқарт-* и т.д.	ET925
в) расширения – сужения	*Соз-, тарылт-, кеңейт, жуандат, жіңішкерт-...*	ET926
Семантический класс: 9. Каузативные глаголы *(продолжение)*		
г) углубления – обмеления.	*Тереңдет-, таяздат-....*	ET927
Временная характеристика: а) продления срока	*Соз-, жылжыт-, ұзарт, қысқарт-* и т.д.	ET928
б) уменьшения срока	*Жылдамдат-, тездет-..*	ET929
Локальная характеристика: а) совместность – разноместность	*Жина-, жинақта-, шашырат-, бөліп отырғыз-*	ET930
б) определенной локализации объектов	*Орнат-, орналастыр-, іл-, тұрғыз-* и т.д.	ET931
Парадигматический ряд глагольных лексем: чистить, очищать, мыть, стирать и т.д.	*Тазала-, тазарт-, тазартқыз-, жу-, жуын-, жуыс-, жуындыр-* и т.д.	ET932
Глаголы, которые имеют дифференциальный признак: изменяя, переводить в другое состояние	*Балқыт-, ез-, ерітт-, қорыт-, жасыт-, жылыт-, ысыт-, қатыр-....*	ET933
Глаголы со значением каузации контакта предметов: соединять, присоединять: а) образуется новый объект	*Жалға-, жалғастыр-, тірке-, ұластыр-, біріктір-, қос-* и т.д.	ET934
б) новый объект не образуется	*тік-, астарла-, тепші-*	ET935
Глаголы со значением каузации функтивных отношений: а) делать готовым к функционированию, использованию	*Дайында-, әзірле-, даярла-, сайла-* и др.	ET936
б) делать что-либо вновь пригодным для употребления	*Жөнде-, түзе-, дұрыс-та-, бүтінде-, өңде-.....*	ET937
в) обеспечивать функционирование объектов	*Күзет-, қарауылда-, бақыла-, тексер-* и др.	ET938
Глаголы со значением каузации движения: 1) нейтральные в смысле направленности движения	*Ағыз-, қозғалт-, көшір-, секірт-, қыдырт-, қадамдат-, саяхатта-* и др.	ET939
2) обозначающие движения по горизонтали и по вертикали	*Бар-, кел-, кет-, қайт-, шық-, кір-, мін-, түс-...*	ET940

Окончание табл. 4

1	2	3
3) обозначающие преодоление преграды в движении	*Ас-, кеш-, өт-, қарғы-* и т.д.	ET941
4) обозначающие темп движения:	*Қу-, айда-,қалқы-, шап-* и т.д.	ET942
- каузация наличия материальных ресурсов (оборудования, вооружения, технических средств)	*Жабдықта-, өбзелде-, безендір-, сайла-, газдандыр-, электрлендір-* и т.д.	ET943
- каузация нематериальных объектов (знания, навыки и т.д.)	*Оқыт-, үйрет-, дайында-, даярла-, білгіз-* и др.	ET944

В последней графе *табл. 4* приведены коды подклассов, которые используются в модели порождения казахского текста.

3.4. Семантические классы причастий *(есімше)* и деепричастий *(көсемше)* казахского языка

Причастия и деепричастия являются особыми формами глагола и представляют собой своеобразный разряд слов, поэтому при их классификации на лексико-семантические группы были использованы, в основном, те же принципы, которые принимались при исследовании глаголов. В казахском языке причастие *(есімше)* представляет собой одну из самых употребительных форм глагола. Роль причастия в казахском предложении более значительна, чем в русском. В казахском языке оно выступает, главным образом, в качестве сказуемого и определения, реже, субстантивируясь, как подлежащее и дополнение. В отличие от русского, причастие в казахском языке является спрягаемой формой глагола. При этом показателем определенного времени являются аффиксы причастия. В отличие от обычного спряжения глаголов, причастия спрягаются как именные слова.

Причастие сохраняет все значения, имеющиеся у глагола. В них различаются переходность и непереходность, они управляют теми же падежами, что и глагол. Причастие может сохранять все залоговые значения того глагола, от которого оно образовано.

Вкратце остановимся на аффиксальных способах образования причастия и деепричастия, так как мы в дальнейших исследованиях будем часто ссылаться на них. Известно, что по способу образования причастия казахского языка делятся на три типа [327, с.301-305]: 1) причастия прошедшего времени с аффиксами на ***-ған/-ген/-қан/-кен;*** 2) причастия будущего времени на ***-ар/-ер/-р;*** 3) причастия переходного прошедшего времени на ***-атын/-етін/-йтын/-йтін***. По своему значению причастие выражает оттенок длительности действия, т.е. действие, которое совершается обычно, часто повторяется, которое происходит в настоящем, произошло

в прошлом и длится продолжительное время. В сложных глаголах основное лексическое значение заключается в первых глагольных компонентах, которые находятся в форме деепричастия, а вторые компоненты – вспомогательные глаголы принимают аффиксы причастия. И здесь второй компонент, т.е. причастие, указывает на незавершенность, многократность и повторямость или, наоборот, завершенность, мгновенность и однократность.

Деепричастия *(көсемше),* так же, как и причастия *(есімше),* составляют особый разряд отглагольных форм, обозначают второстепенные действия и служат признаком другого действия. Как сказано в Академической грамматике "Қазіргі қазақ тілі": "Одна из основных черт деепричастий состоит в том, что деепричастия в казахском языке, а также и в других тюркских языках, выражают действия "побочные", сопровождающие "главное" [163, с.320].

Как пишет Н. Г. Сауранбаев: "Надо сказать, что деепричастия не заменяют русских приставок. Сходные черты их состоят в том, что они внешне выступают как словообразующие элементы. Но в семантическом отношении они расходятся... Казахские деепричастия выражают собой специфические виды казахских глаголов – начало, конец процесса действия" [250, с.48].

Деепричастия, выражающие незаконченное действие, происходящее одновременно с действием основного глагола, и указывающие условия совершения действия, имеют аффиксы: ***-а/-е/-й*** *(айт+а, тер+е, сөйле+й* и т.д.*),* а выражающие действие однократное, предшествующее главному действию и указывающее, при каких обстоятельствах совершается действие, выраженное основным глаголом ***-ып/-іп/-п*** *(айт+ып, жүр+іп, кел+іп, ал+ып, қара+п, ойла+п* и т.д.*).* И наконец, деепричастия, выражающие действие предполагаемое, являющееся целью, а также указывающее на время действия, выраженного спрягаемой формой глагола, образуются от корня или основы глагола путем присоединения аффиксов: ***-қалы/ -ғалы/-келі/-гелі*** *(айт+қалы, ек+келі, кет+келі, сөйле+гелі, апар+ғалы, жаз+ғалы* и т.д.). Присоединение к деепричастиям аффиксов ***-ыңқыра/-іңкіре*** снижает темп, эффективность действия *(қорқ+ыңқыра - побаиваясь, іш+іңкіре - допивая, қызар+ыңқыра - покраснев* и т.д.*).* Аффиксы ***-ғыла/-гіле/-қыла/-кіле*** в соединении с деепричастыми аффиксами выражают многократность действия *(жұл+қыла+п - подергивая, бар+ғыла+п - похаживая, тер+гіле+п - собирая* и т.д.*).*

Перечисленные и другие признаки и способы образования причастий и деепричастий использованы в нашем исследовании при выделении их семантических классов *(табл. 5 и 6).*

Таблица 5. **Семантические классы причастий *(есімше)*, использованных в модели порождения казахского текста**

Семантический подкласс причастий	Прошед. вр. ***(-ған, -ген, -қан, -кен)***	Код под-класса	Будущее время ***(-ар/-ер/-р)***	Код под-класса	Будущее время ***(-атын, -етін, -йтын, -йтін)***	Код под-класса
1	2	3	4	5	6	7
Семантический класс:	**1. Причастия, обозначающие признаки физических действий**					
1. Свойственные человеку, особям животного мира и элементам неживой природы	*Шулаған, азайған, тозған ...*	ЕШ001	*Шулар, азайар, тозар* и т.д.	ЕШ041	*Шулайтын, азайатын, тозатын ...*	ЕШ081
2. Физическо-психических и эмоциональных действий, свойственных только человеку	*Іренжіген, урысқан, қайғырған...*	ЕШ002	*Іренжір, урысар, қайғырар ...*	ЕШ042	*Іренжитін, урысатын, қайғыратын ...*	ЕШ082
3. Специфичные для элементов неживой природы	*Жауған, аққан (су), күркіреген ...*	ЕШ003	*Жауар, ағар, күркірер, сарқырар ...*	ЕШ043	*Жауатын, күркірейтін, ұйтқыйтын ...*	ЕШ083
4. Свойственные только человеку	*Сіңбірген, сатқан ...*	ЕШ004	*Сіңбірер, сатар* и т.д.	ЕШ044	*Сіңбіретін, сататын* и т.д.	ЕШ084
5. Свойственные человеку и особям животного мира	*Жүгірген, ішкен, жеген* и т.д.	ЕШ005	*Жүгірер, ішер, жер, шаршар...*	ЕШ045	*Жүгіретін, ішетін, жейтін* и т.д.	ЕШ085
Семантический класс:	**2. Причастия, обозначающие признаки физических состояний**					
1. Свойственные человеку, особям животного мира и элементам неживой природы	*Тозған, толысқан, сақталған...*	ЕШ011	*Тозар, толысар, сақталар...*	ЕШ051	*Тозатын, толысатын, сақталатын ...*	ЕШ091
2. Свойственные элементам неживой природы	*Көктеген, піскен, үйілген, жауған*	ЕШ012	*Көктер, пісер,үйілер, жауар ...*	ЕШ052	*Көктейтін, пісетін, үйілетін*	ЕШ092

Продолжение табл. 5

1	2	3	4	5	6	7
3. Свойственные человеку и элементам неживой природы	*Жайнаған, күлімдеген* и т.д.	ЕШ013	*Жайнар, жарқырар, күлімдер ...*	ЕШ053	*Жайнайтын, күлімдейтін* и т.д.	ЕШ093
4. Свойственные только человеку	*Байыған, кедейленген, жауласқан ...*	ЕШ014	*Байыр, кедейленер, жауласар ...*	ЕШ054	*Байыйтын, кедейленетін жауласатын*	ЕШ094
Семантический класс:	**3. Причастия обозначающие признаки психических действий**					
1. Признаки психических и эмоционально-психических действий, свойственных только человеку	*Күйінген, толқыған, сүйсінген ...*	ЕШ021	*Күйінер, толқынар, сүйсінер ...*	ЕШ061	*Күйінетін, толқитын, сүйсінетін...*	ЕШ101
2. Признаки психических действия-мышления	*Ойланған, сөйлеген, айтысқан ...*	ЕШ022	*Ойланар, сөйлер, айтысар ...*	ЕШ062	*Ойланатын, сөйлейтін, айтысатын*	ЕШ102
3. Признаки психических действий, приписываемых элементам неживой природы	*Тұнжыраған, күлімдеген, мұңайған ...*	ЕШ023	*Тұнжырар, күлімдер, қайғырар, мұңайар ...*	ЕШ063	*Тұнжырайтын, күлімдейтін, мұңайатын ...*	ЕШ103
4. Свойственные человеку, особям животного мира и приписываемые элемамм неживой природы	*Қайғырған, қуанған, жабырқаған ...*	ЕШ024	*Қайғырар, қуанар, жабырқар ...*	ЕШ064	*Қайғыратын, қуанатын* и т.д.	ЕШ104
Семантический класс:	**4. Причастия, обозначающие признаки психических состояний**					
1. Признаки психических и эмоционально-психических состояний, специфичных для человека	*Өкпелеген, сүйсінген, күрсінген ...*	ЕШ031	*Өкпелер, сүйсінер, күрсінер ...*	ЕШ071	*Өкпелейтін, сүйсінетін, күрсінетін ...*	ЕШ111
2. Специфичные человеку, особям животного мира и приписываемые элементам	*Долданған, мұңайған, қажыған,*	ЕШ032	*Долданар, мұңайар, қажыр,*	ЕШ072	*Долданатын, мұңайатын, қажыйтын,*	ЕШ112

Окончание табл. 5

1	2	3	4	5	6	7
4. Обозначающие признаки психического состояния-мышления, свойственные человеку	*Ойланған, тұжырымдаған, пікірлескен* и т.д.	ЕШ034	*Ойланар, қортындылар, тұжырымдар* ...	ЕШ074	*Ойланатын, қортындылайтын, тұжырымдайтын* и т.д.	ЕШ114
5. Признаки психических состояний, свойственные человеку и особям животного мира	*Мұңайған, қажыған, боздаған, қайғырған, қамыққан* и т.д.	ЕШ035	*Мұңайар, қажыр, боздар, қайғырар, қамығар* и т.д.	ЕШ075	*Мұңайатын, қажыйтын, боздайтын, қайғыратын, қамығатын* и т.д.	ЕШ115

В графах 3, 5 и 7 *табл. 5* приведены коды семантических подклассов причастий, которые будут использованы в алгоритме программы порождения казахского текста. ЕШ – сокращение от слова *есімше.*

Таблица 6. **Семантические классы деепричастий *(көсемше)*, использованных в модели порождения казахского текста**

Семантические подклассы деепричастий	Примеры и коды деепричастия					
	Незаконченное действие (с аффикс. на ***-а/-е/-й***)	Код под-класса	Однократное действие (с аффикс. на ***-ып/-іп/-п***)	Код ?од-класса	Предполагаемое действие (аф. на ***-қалы/ -галы,-келі/-гелі***)	Код под-класса
1	2	3	4	5	6	7
Семантический класс:	**1. Деепричастия, обозначающие физические действия**					
1. Свойственные человеку, особям животного мира и приписываемые элементам неживой природы	*Жата, тұра, жүре, шулай* и т.д.	КШ001	*Жатып, тұрып, жүріп, шулап* и т.д.	КШ041	*Жатқалы, тұрғалы, жүргелі, шулағалы* и т.д.	КШ081
2. Приписываемые элементам неживой природы	*Жата, тұра, шулай, күркірей, әндете*	КШ002	*Жатып, тұрып, шулап, күркіреп, әндетіп* и т.д.	КШ042	*Жатқалы, тұрғалы, шулағалы, әндеткелі* и т.д.	КШ082
3. Обозначающие физические действия, свойственные только человеку	*Жуа, пісіре, кигізе, сіңбіре, шымши ...*	КШ003	*Жуып, пісіріп, кигізіп, шымшып* и т.д.	КШ043	*Жуғалы, пісіргелі, кигізгелі* и т.д.	КШ083
4. Свойственные человеку и особям животного мира	*Жүгіре, жарыса, секіре, іше, жей ...*	КШ004	*Жүгіріп, жарысып, секіріп, ішіп, жеп*	КШ044	*Жүгіргелі, жарысқалы, секіргелі* и т.д.	КШ084
5. Физические действия движения, свойственные человеку, особям животного мира и неживой природе	*Қозғала, жылжи, ұша, секіре, домалай ...*	КШ005	*Қозғалып, жылжып, ұшып, секіріп, домалап ...*	КШ045	*Қозғалғалы, жылжығалы, секіргелі, ұшқалы* и т.д	КШ085
6. Физические эмоциональные действия, свойственные человеку	*Күле, ыржиа, шапалақтай, сүйе, алақайлай* и т.д.	КШ006	*Күліп, ыржыйып, шапалақтап, сүйіп, алақайлап* и т.д.	КШ046	*Күлгелі, ыржыйғалы, шапалақтағалы, сүйгелі*	КШ086

Продолжение табл. 6

1	2	3	4	5	6	7
Семантический класс:	**2. Деепричастия, обозначающие физические состояния**					
1. Обозначающие физические состояния, свойственные только человеку	*Жеңілдене, шыныға, рахаттана...*	КШ011	*Жеңілденіп, шығып, рахаттанып*	КШ051	*Жеңілдегелі, шыққалы, рахаттанғалы*	КШ091
2. Свойственные человеку, особям животного мира и приписываемые элементам неживой природы	*Абыржый, тыныштала, беймазалана, жарқырай ...*	КШ012	*Абыржып, тыныштальп, беймазаланып* и т.д.	КШ052	*Абыржығалы, тыныштальғалы, беймазаланғалы* и т.д.	КШ092
3. Физические состояния, приписываемые элементам неживой природы	*Абыржый, тыныштала, жадырай, күңірене* и т.д.	КШ013	*Абыржып, тыныштальп, жадырап, күңіреніп* и т.д.	КШ053	*Абыржығалы, тыныштальғалы, жадырағалы* и т.д.	КШ093
4. Физические состояния, свойственные особям животного мира	*Абыржый, тыныштала, беймазалана* и т.д.	КШ014	*Абыржып, тыныштальп, беймазаланып* и т.д.	КШ054	*Абыржығалы, тыныштальғалы, беймазаланғалы ...*	КШ094
5. Физические состояния, свойственные человеку и элемен-м неживой природы	*Абыржый, тыныштала, бусана* и т.д.	КШ015	*Абыржып, тыныштальп, бусанып* и т.д.	КШ055	*Абыржығалы, тыныштальғалы* и т.д.	КШ095
Семантический класс:	**3. Деепричастия, обозначающие психические действия**					
1. Психические действия-мышления	*Қортындылай, тұжырымдай, ойлай, сөйлей*	КШ021	*Қортындылап, тұжырымдап, ойлап, сөйлеп*	КШ061	*Қортындылағалы, тұжырымдағалы ...*	КШ101
2. Психо-физические действия, свойственные только человеку	*Ашулана, қуана, ұрыса, шапалақтай*	КШ022	*Ашуланып, қуанып, ұрысып, қуантып*	КШ062	*Ашуланғалы, ұрысқалы, қуанғалы* и т.д.	КШ102
3. Психичические действия ощущений, желаний, свойственных человеку	*Жирене, сүйсіне, тұшына, сағына* и т.д.	КШ023	*Жиреніп, сүйсініп, тұшынып, сағынып*	КШ063	*Жиренгелі, сүйсінгелі, сағынғалы* и т.д.	КШ103

Окончание табл. 6

1	2	3	4	5	6	7
4. Психичические действия внимания, свойственные только человеку	*Үйрете, көңілдене, сүйсіне* и т.д.	КШ024	*Үйретіп, көңілденіп, сүйсініп* и т.д.	КШ064	*Үйреткелі, көңілденгелі, сүйсінгелі* и т.д.	КШ104
5. Психические активные действия, свойственные человеку и особям животного мира	*Бақыра, жұлына, долдана, жұлқына, дүрсілдей* и т.д.	КШ025	*Бақырып, жұлынып, долданып, дүрсілдеп* и т.д.	КШ065	*Бақырғалы, жұлынғалы, долданғалы, жұлқынғалы*	КШ105
Семантический класс:	**4. Деепричастия, обозначающие психические состояния**					
Психические состояния, свойственные человеку	*Қуана, жадырай, мұңая, масайрай* и т.д.	КШ031	*Қуанып, жадырап, мұңайып, масайрап* и т.д.	КШ071	*Қуанғалы, жадырағалы, мұңайғалы, масайрағалы*	КШ111

В графах 3, 5 и 7 *табл. 6* приведены коды семантических подклассов деепричастий, которые будут использованы в модели порождения казахского текста. КШ – сокращение от слова *көсемше.*

3.5. Семантические классы прилагательных *(сын есім)* казахского языка

Специфические формы и семантические особенности имени прилагательного в казахском языке четко отграничивают его от других частей речи. Например, признаками, обособляющими и отграничивающими его от имен существительных являются такие признаки, как отсутствия формантов, несклоняемость, отсутствие множественного числа, не принятие аффиксов притяжательности. Одна из особенностей казахского прилагательного – наличие морфологических форм, т.е. степени сравнения, а также его препозитивность. Последнее выражается в том, что казахские прилагательные могут стоять только перед определяющим словом (за исключением личных имен) [107]. Постпозиция прилагательного может встречаться в поэтической речи или самостоятельном предложении (*Күн жылы – День теплый. Ай жарық – Луна светлая*).

Исследователи различных языков почти единогласно делят прилагательные на качественные и относительные. Однако выделение внутри этих классов более мелких лексико-семантических групп проводится неоднозначно, с использованием самых различных принципов и критериев [12, с.211-215; 64, с.9-28; 318]. Отмечается, что качественные прилагательные крайне неоднородны, так как они в основном освобождены от предметного значения (абстрагированы). Как правило, их дифференцируют по их роли в высказывании, в то время как относительные прилагательные могут иметь любые значения, связанные с предметами и событиями. Все это вынуждает нас искать такую семантическую классификацию прилагательных, которая была бы удобна для цели порождения казахского текста и в то же время была относительно небольшой по числу семантических классов. Предлагаемая нами семантическая классификация прилагательных опирается на денотативно-десигнативный принцип. Она, естественно, не охватывает всех возможных типов, но вполне достаточна для решения задач нашего исследования. Разработанные нами классификационные таблицы качественных и относительных казахских прилагательных приведены в *табл. 7 и 8.*

Таблица 7. **Семантические классы качественных прилагательных *(сапалық сын есімдер)*, используемые в модели порождения казахского текста**

Семантический подкласс прилагательного	Положительная степень *(жай шырай)*	Код под-класса	Сравнительная степень *(салыстырмалы шырай)*	Код под-класса	Превосходная степень *(күшейтпелі шырай)*	Код под-класса
1	2	3	4	5	6	7
Семантический класс:	**1. Качественные прилагательные, обозначающие признаки, свойственных человеку и тому, что связано с ним**					
1. Означающие внутреннее качество человека	*Жақсы,жаман,тәуір, алғыр, жүйрік* и т.д.	CC001	*Жақсырақ, жамандау, жүйрігірек ...*	CC041	*Өте жақсы, тіпті жаман, тым салақ* и т.д.	CC091
2. Означающие физиологические свойства человека	*Семіз, арық, шолақ, ақсақ, соқыр* и т.д.	CC002	*Семіздеу, арықтау, соқырлау* и т.д.	CC042	*Өте семіз, тым арық, тым соқыр* и т.д.	CC092
3. Означающие поведение человека	*Жуас, тентек, пысық, қыңыр, жұмсақ* и т.д.	CC003	*Жуастау, тентектеу, жайдарырақ*	CC043	*Өте жуас, тым тентек, өте пысық, өте қорқақ* и т.д.	CC093
4. Употребляющиеся в связи с температурными изменениями окружающей среды	*Қызу, ыстық, жылы, суық, салқын, дымқыл, құрғақ...*	CC004	*Қызулау, ыстықтау, жылылау, суығырақ*	CC044	*Аса қызу, тым ыстық, өте жылы, тым суық, тым салқын* и т.д.	CC094
5. Употребляющиеся в связи с возрастом человека	*Қарт, жас, кәрі, балғын, сәби* и т.д.	CC005	*Қарттау, жастау, кәрілеу* и т.д.	CC045	*Тым қарт, өте жас, тым кәрі, тым сәби* и т.д.	CC095
6. Употребляющиеся в связи с бытовой жизнью человека	*Бай, кедей, жарлы* и т.д	CC006	*Байырақ, кедейлеу, жарлылау* и т.д.	CC046	*Өте бай, өте кедей,тым кедей, тым жарлы* и т.д.	CC096

Продолжение табл. 7

1	2	3	4	5	6	7
7. Употребляющиеся в связи с внешней красотой человека	*Сұлу, әсем, әдемі, келісті, көркем* и т.д.	CC007	*Сұлуырақ, сұлулау, әсемірек, әдемілеу*	CC047	*Тым сұлу, өте сұлу, өте әсем, тым әдемі* и т.д.	CC097
Семантический класс:	**2. Качественные прилагательные, обозначающие признаки, свойственые особям животного мира**					
1. Означающие физиологическое свойство животных	*Семіз, арық, жүдеу, ақсақ, тапал, қортық*	CC011	*Семіздеу, арықтау, биігірек, ақсақтау*	CC051	*Аса сұлу, тым арық, өте жүдеу, тым ақсақ* и т.д.	CC101
2. Означающие поведение животных	*Асау, жуас, қорқақ, қыңыр тентек*	CC012	*Асаулау, жуастау, қорқақтау ...*	CC052	*Тым асау, өте жуас, тым қорқақ, өте қыңыр* и т.д.	CC102
3. Употребляющиеся в связи с возрастом животных	*Кәрі, жас* и т.д.	CC013	*Кәріеу, жастау* и т.д.	CC053	*Тым кәрі, өте кәрі, тым жас, өте жас* и т.д.	CC103
4. Употребляющиеся в связи с внешним видом животных	*Торы, бурыл, ақ, боз, қара, шұбар, таргыл*	CC014	*Торылау, бурылдау, ақтау, боздау*	CC054	*Тым торы, тым бурыл, аппақ, қап-қара* и т.д.	CC104
Семантический класс:	**3. Качественные прилагательные, обозначающие признаки, используемые для описания предметов**					
1. Означающие цвета предметов	*Ақ, көк,қызыл, сары, сұр, боз, шұбар, жасыл*	CC021	*Ақшыл, көкшіл, қызылырақ, қызылдау*	CC061	*Аппақ, көкпеңбек, қып-қызыл, сап-сары, сұп-сұр* и т.д.	CC111
2. Означающие вкусовые качества пищевых продуктов	*Тәтті, ащы, тұщы, қышқыл* и т.д	CC022	*Тәттілеу, ащылау, ащырақ, қышқылдау...*	CC062	*Өте тәтті, тым ащы, аса дәмді* и т.д.	CC112

Продолжение табл. 7

1	2	3	4	5	6	7
3. Означающие объем предметов	*Дәу, ірі, үлкен, кіші, қалың, жуқа, жуан...*	CC023	*Ірілеу, үлкенірек, үлкендеу, кішірек* и т.д.	CC063	*Тым ірі, өте дәу, тым кіші, өте қалың, өте жуқа* и т.д.	CC113
4. Означающие вес предмета	*Ауыр, жеңіл, зіл* и т.д	CC024	*Ауырлау, жеңілдеу, жеңілірек, зілдей...*	CC064	*Өте ауыр, тым ауыр, өте жеңіл, тым жеңіл*	CC114
5. Означающие природное состояние предмета	*Қалың, сирек, жуан, жіңішке, селдір, қаба*	CC025	*Қалыңдау, сирегірек, жуандау, селдірлеу*	CC065	*Қап-қалың, тым сирек, өте жуан, сеп-селдір* и т.д.	CC115
6. Означающие признак предмета в связи с температурными изменениями окружающей среды	*Ыстық, суық, салқын, жылы, дымқыл, құрғақ* и т.д.	CC026	*Ыстықтау, суығырақ, салқындау, құргақтау. . .*	CC066	*Ып-ыстық, тым ыстық, сүп-суық, тым суық, сап-салқын* и т.д.	CC116
7. Употребляющиеся в связи с внешней формой предмета	*Дөңгелек, сопақ, үшкіл, тегіс, жұмыр ...*	CC027	*Дөңгелектеу, сопақтау, үшкілдеу* и т.д.	CC067	*Дөп-дөңгелек, тым сопақ, өте үшкір, теп-тегіс* и т.д.	CC117
8. Употребляющиеся в связи с природными явлениями	*Жарық, қараңғы, ашық, күңгірт, жылы* и т.д.	CC028	*Жарықтау, қараңғылау, ашығырақ* и т.д.	CC068	*Жап-жарық, тым жарық, өте ашық, тым күңгірт* и т.д.	CC118
9. Употребляющиеся в связи с количественными отношениями предметов	*Аз, көп, мол, шағын, қыруар, кішкене* и т.д.	CC029	*Аздау, көптеу, көбірек, молдау, молырақ* и т.д.	CC069	*Өте аз, тым аз, өте көп, шап-шағын, моп-мол* и т.д.	CC119
10. Употребляющиеся в связи со стоймостью предмета	*Арзан, қымбат, тегін, бос* и т.д.	CC030	*Арзанырақ, қымбаттау* и т.д.	CC070	*Ап-арзан, тым арзан, теп-тегін* и т.д.	CC120

Окончание табл. 7

1	2	3	4	5	6	7
Семантический класс:	**4. Качественные прилагательные, обозначающие оценочные признаки, свойственные человеку, особям животного мира и используемые для описания предмета**					
1. Обозначающие цвета и качества предмета	*Ақ, қара, жақсы, жаман тәуір* и т.д.	CC031	*Ақшыл, қаралау, жақсырақ, жамандау ...*	CC081	*Аппақ, қап-қара, тым қара, өте жақсы, тәптәуір ...*	CC131
2. Обозначающие объем	*Үлкен, кіші, дәу* и т.д	CC032	*Үлкендеу,молырақ, молдау, кішірек* и т.д.	CC082	*Үп-үлкен, өте үлкен, тым кіші, өте дәу* и т.д.	CC132
3. Обозначающие вес	*Ауыр, жеңіл, зіл* и т.д.	CC033	*Ауырлау, жеңілірек, жеңілдеу, зілдей ...*	CC083	*Ап-ауыр, өте ауыр, жеп-жеңіл, тым жеңіл* и т.д.	CC133
4. Употребляющиеся в связи с возрастом живого и неживого	*Қарт, кәрі, жас* и т.д.	CC034	*Қартаңдау, кәрілеу, жасырақ, жастау*	CC084	*Тым қарт, өте кәрі, тым кәрі, жап-жас* и т.д.	CC134

В графе 4 даны примеры качественных прилагательных, относящиеся к сравнительной степени, которые образуются с помощью аффиксов: ***-рақ/-рек, -ырақ/-ірек, -лау/-леу, -дау/-деу, -тау/-теу, -қыл/-ғыл, -қылт/-ғылт, -тым,-шыл/-шіл, -қай, -аң*** и др. В графе 6 – примеры превосходной степени, которые образуется путем сочетания прилагательного с наречиями: *өте, аса, тым, тіпті, тіптен, асқан, ең, нағыз и др.;* путем удвоения начального слога прилагательного. Например *қып-қызыл – прекрасный, наикрасный, сап-сары – прежелтый, наижелтейший, біп-биік – наивысший, превысокий* и т.д.

В графах 3, 5 и 7 *табл. 7* даны условно принятые нами коды семантического подкласса, которые будут использованы в модели порождения казахского текста. CC – сокращение от словосочетания *сапалық сын есім.*

Таблица 8. **Семантические классы относительных прилагательных *(қатыстық сын есім)*, используемые в модели порождения казахского текста**

Примеры относительных прилагательных	Код класса	Сравнительная степень *(салыстырмалы шырай)*	Код класса	Превосходная степень *(күшейтпелі шырай)*	Код класса
1	2	3	4	5	6
Семантический класс:	1. Прилагательные, характеризущие объект по его отношению к какому-либо одушевленному субъекту				
Тершең, кіршең, малды, безеулі, сақалды, азулы, қанатты,үйлі, бассыз	ҚС011	*Тершеңдеу, кіршеңірек, малдылау, безеулілеу, сақалдырақ, азулырақ ...*	ҚС012	*Өте тершең, тым кіршең, нағыз малды, тым безеулі, тым сақалды* и т.д.	ҚС013
Семантический класс:	2. Прилагательные, характеризущие объект по его отношению к какому-либо неодушевленному предмету				
Таулы, сулы, желді, күзгі, жазғы, түнгі, күндізгі, өзенді, сулы ...	ҚС021	*Таулылау, сулырақ, күзгілеу, жазғылау, сулырақ, таусыздау, сусыздау* и т.д.	ҚС022	*Ең таулы, тым сулы, тіпті желді, нағыз жазғы, ең таулы-тасты* и т.д.	ҚС023
Семантический класс:	3. Прилагательные, характеризущие объект по его отношению ко времени				
Кешкі, түскі, күзгі, түнгі, жазғы, айлық	ҚС031	*Кешкірек, түскірек, түнгілеу, жазғылау* и т.д.	ҚС032	*Тым кешкі, нағыз түскі, нағыз жазғы* и т.д.	ҚС033
Семантический класс:	4. Прилагательные, характеризущие объект по его отношению к материалу				
Көйлектік, пальтолық, тағамдық, шайлық, соғымдық и т.д.	ҚС041	*Көйлектіктеу, пальтолықтау, едендіктеу, соғымдықтау* и т.д.	ҚС042	*Нағыз көйлектік, нағыз пальтолық, ең тағамдық, нағыз едендік* и т.д.	ҚС043
Семантический класс:	5. Прилагательные, характеризущие объект по его отношению к какому-либо действию				
Жылауық, сыбырлауық, асыранды, жаттанды, жасырынды, ауыспалы	ҚС051	*Жылауықтау, жаттандылау, ауыспалылау, жылжымалылау, тапқыштау*	ҚС052	*Нағыз жылауық, тіпті сыбырлауық, ең жаттанды, тым жасырынды* и т.д.	ҚС053

Окончание табл. 8

1	2	3	4	5	6
Семантический класс:	**6. Прилагательные, характеризующие объект по его отношению к месту**				
Жерлес, ауылдас, көршілес, елдес, ұялас, іргелес, мекендес и т.д.	ҚС061	*Жерлесшіл, ауылдасшыл, көршілестеу, ұяластау, іргелестеу, мекендестеу*	ҚС062	*Нағыз жерлес, нағыз ауылдас, тіпті көршілес, нағыз ұялас, өте іргелес* и т.д.	ҚС063
Семантический класс:	7. Прилагательные, характеризующие объект по его отношению к отвлеченному понятию				
Аршыл, ойшыл, сезімтал, түйсіксіз, ақылсыз, еріксіз, санасыз, пікірлес, мұңдас, ниеттес, сырлас и т.д.	ҚС071	*Аршылдау, ойшылдау, сезімталырақ, түйсіксіздеу, ақылсыздау, санасыздау, пікірлестеу, мұңдасырақ, ниеттестеу* и т.д	ҚС072	*Аса аршыл, өте ойшыл, тым сезімтал, нағыз түйсіксіз, тіптен ақылсыз, өте санасыз, нағыз пікірлес, өте сырлас* и т.д.	ҚС073

В графе 3 даны примеры относительных прилагательных на сравнительную степень, которые образуются с помощью аффиксов: ***-рақ/-рек, -ырақ/-ірек, -лау/-леу, -дау/-деу, -тау/-теу, -қыл/-ғыл, -қылт/-ғылт, -тым, -шыл/-шіл, –қай, -аң*** и др.

В графе 5 – на превосходную степень, образованные путем сочетания прилагательного с наречиями: *өте, аса, тым, тіпті, тіптен, асқан, ең, нағыз и др.*

В графах: 2, 4 и 6 даны разработанные нами коды семантических подклассов относительных прилагательных. ҚС – сокращение от словосочетания – *қатыстық сын есім.*

3.6. Семантическая классификация наречий *(үстеу)*

Известно, что казахские слова-наречия по своему значению не однородны, так как одно и то же наречие в зависимости от контекста может иметь разное значение. При семантической классификации таких слов в казахском языке принято слова, относящиеся к наречиям, делить на следующие две основные группы [164, с.203]:

1) определительные наречия *(анықтауыш үстеулер),* которые выражают качество, количество и меру, образ и манеру действия, способ и совместность действия;

2) обстоятельственные наречия *(пысықтауыш үстеулер),* которые указывают на время, место, причину и цель действия.

В казахском языке качественные наречия не выделяются в особую группу, но при этом ряд наречий обладает значением качественных наречий русского языка (*медленно – шабан, бәсең, баяу; быстро – тез, шапшаң* и т.д.) [276, с.340].

По нашим статистическим наблюдениям в современном казахском языке по степени употребительности чаще остальных стали встречаться наречия образа и способа действия, затем наречия времени, места, количества и меры, менее чаще – наречия с оттенком усиления, выражая качество предмета, самого действия, кроме того, меры или объема как с оттенком усиления, так и ослабления. В группе наречий образа и способа совершения действия наряду с наречиями степени усиления и ослабления действия в отдельную семантическую группу можно отнести и наречия совокупности *(топтау үстеулері)* [276, с.338].

Беря за основу названные две группы и используя известные в литературе принципы классификации наречий (и шире – слов категории состояния[1]), а также данные, полученные по нашим экспериментальным текстам, мы разработали приемлемую для нас семантическую классификацию наречий казахского языка, которая и будет использована в построении модели порождения казахского текста *(табл. 9).*

[1] В казахском языке категория состояния грамматически не выражается [276, с.353].

Таблица 9. **Семантические классы наречий, использованные в модели порождения текста**

Семантический подкласс наречия	Примеры подкласса наречий	Код подкласса
Семантический класс: 1. Определительные наречия *(Анықтауыш үстеулер)*		
1	2	3
1. Определительные наречия образа действия *(бейне, сын үстеулері)*	*Осылай, осылайша, сөйтіп, өйтіп, бүйтіп, олай-бұлай, бірден, бірдей, бірге, бірте-бірте, біртіндеп, ойша, өзінше, іле, қолма-қол, кезек-кезек, көрнеу, өздігінен, өзді-өзді, жатпай-тұрмай, үнемі* и т.д.	YC011
2. Определительные наречия количества и меры *(мөлшер үстеулері)*	*Онша, сонша, осынша, оншалық, соншалық, осыншалық, оншама, мұншама, анағұрлым, бірен-саран, ондаған, көптеген, бірталай* и т.д.	YC012
3. Определительные наречия качества *(сапалық үстеулері)*: а) определительные наречия с оттенком усиления или ослабения *(күшейту не әлсірету үстеулері)*	*Ең, әбден, ылғи, кілең, сәл, өңкей, тіпті, тым, нақ, нағыз, әнтек, мүлдем, дәл, керемет, қабағат, мейлінше, сонша, төтенше, жөнсіз, орасан, өрен, аса, өте* и т.д.	YC013
б) определительные наречия с оттенком собирательности или разделительности *(Топтау, саралау үстеулері)*	*Екеулеп, үшеулеп, он-ондап, бір-бірлеп, он-оннан, қырық-елуден, аз-аздан, көп-көптен, топ-тобымен, рет-ретімен, бас-басына, үйді-үйіне* и т.д.	YC014
Семантический класс: 2. Обстоятельственные наречия *(Пысықтауыш үстеулер)*		
1. Обстоятельственные наречия времени *(мезгіл үстеулері)*	*Бүгін, былтыр, ертең, таңертең, кешке, қазір, енді, әлі, ендігәрі, екіншәрі, бұрын, әуел-баста, ерте, кеш, күндіз, бүрсігүні, әсте, күнімен, ертеден, күні бүгін, күні кешке, бүгін таңда, күн ара, күні бойы, қыстыгүні, қыстай, жаздай, ала жаздай, күні-түні, анда-санда, енді-енді, осындайда, апақ-сапақта* и т.д.	YC021
2. Обстоятельственные наречия причины *(себеп-салдар үстеулері)*	*Жоққа, босқа, құр босқа, бекерге, амалсыздан, лажсыздан, шарасыздан, сөйле-сөйлей, көре-көре, оқый-оқый* и т.д.	YC022
3. Обстоятельственные наречия цели *(мақсат үстеулері)*	*Әдейі, әдейілеп, жорта, қасақана, кітапқа бола, оқуға бола* и т.д.	YC023
4. Обстоятельственные наречия места *(мекен үстеулері)*	*Ілгері, ілгеріде, әрі, әріде, әрмен, жоғары, төмен, жолшыбай, жол-жөнекей, алға, артта, алды-артынан, тысқары, одан-будан* и т.д.	YC024

Примеры покласса наречий заимствованы из книги А. Ысқақов. Қазіргі қазақ тілі. Алматы: Ана тілі, 1991. – 340-343-бб.

3.7. Семантическая классификация местоимении *(есімдік)*

По семантическим свойствам местоимения в языкознании относят к дейктическим словам. Самыми распространенными в казахском языке являются личные *(жіктеу есімдігі)* и указательные *(сілтеу есімдігі)* местоимения. В отличие от существительных, прилагательных и числительных, местоимения в языке имеют максимально общее и вместе с тем всеобъемлющее значение и выполняют указательную функцию: "обозначая предметы или признаки, они не называют и не характеризуют отдельные конкретные предметы, признаки или определенные количества, а лишь указывают на них" [277]. Известно, что когда в предложении существительное (или другое слово) заменено местоимением, то это местоимение не несет в себе все значения заменяемого слова. Местоимение только указывает на одно из его значений или стремится к этому. Значение же местоимения при этом является очень обобщенным и зависит в основном от развития ситуации, не обладает объективностью, "привязано" к определенному субъекту, о котором идет повествование в тексте. Также известно, что употребление местоимений тесно связано с употреблением имен и является как бы их синонимами. Тем самым, автор текста старается избежать частого повтора одного и того же слова. Таким образом, денотативное и сигнификативное значении местоимения зависят от очень большого числа факторов.

В казахском языке местоимения немногочисленны. Об этом говорят следующие факты. А. Исхаков в грамматике [327, с.210] отмечает, что число местоимений в казахском языке не более 60–70, а, по данным И. К. Уйыкбаева, это число равно 111 [290]. Наши статистические подсчеты на компьютере по тексту романа М.Ауэзова "Абай жолы" показали, что количество местоимений достигает до 150 (вместе с заимственными и диалектными формами). В газетных текстах местоимений 102, а по "Словарю языка Абая" - 68 [14]. Такая малочисленность слов этой категории дает нам возможность полностью охватить их при сематической классификации. Как известно, по значению местоимения разделяются на следующие семь разрядов: 1) личные, 2) указательные, 3) вопросительно-относительные, 4) неопределеные, 5) возвратные, 6) обобщительные, 7) отрицательные. Заметим, что в казахской грамматике, так называемые, "притяжательные местоимения" не фигурируют в качестве самостоятельного разряда, а определительные местоимения выступают под названием "обобщающие местоимения".

Местоимения являются одним из основных средств связи предложений в абзаце и абзацев в единый текст. Помимо этого, как это будет показано, личные местоимения выступают, как правило, в роли контекстных синонимов главных и второстепенных опорных слов текста. Учитывая отмеченную значимость местоимений в структуре текста и тот факт, что общее число местоимений в казахском языке относительно не велико, в таблице семантических классов местоимений *(таблица 10)* мы вводим для каждого местоимения индивидуальный семантический код. В ней так же приведены общеупотребительные классы местоимений.

Таблица 10. **Семантические классы местоимений, использованных в модели порождения текста**

Семантический класс местоимения	Конкретное местоимение класса	Код место-имения
1	2	3
1. Личные местоимения *(жіктеу есімдіктері)*	*Мен*	ЕС101
	Біз	ЕС102
	Ол	ЕС103
	Сен	ЕС104
	Сіз	ЕС105
	Біздер	ЕС106
	Олар	ЕС107
	Сіздер	ЕС108
	Сендер	ЕС109
2. Возвратные местоимения *(өздік есімдіктері)*	*Өз*	ЕС201
	Өзі, өзім, өзің, өзіңіз, өзіміз, өздеріміз, өздерің, өздеріңіз, өздері и т.д.	ЕС 202-С299
3. Указательные местоимения *(сілтеу есімдіктері):* а) местоимения, указывающие на предметы и явления, находящиеся относительно ближе от говорящего	*Осы*	ЕС301
	Бұл	ЕС302
	Мына	ЕС303
	Мынау	ЕС304
	Осынау	ЕС305
	Міне	ЕС306
	Әне	ЕС307
б) местоимения, указывающие на предметы и явления, находящиеся относительно дальше от говорящего	*Ол*	ЕС311
	Сол	ЕС312
	Ана	ЕС313
	Анау	ЕС314
	Сонау	ЕС315
	Сона	ЕС316
	Түу	ЕС317
4. Обобщающие местоимения (определительные местоимения) *(жалпылау есімдіктері)*	*Әр*	ЕС401
	Әрбір	ЕС402
	Әркім	ЕС403
	Ең	ЕС404
	Бар	ЕС405
	Барлық	ЕС406
	Барша	ЕС407
	Бәрі	ЕС408
	Бүкіл	ЕС409
	Бүтін	ЕС410
	Күллі	ЕС411
	Түгел	ЕС412
5. Вопросительно-относительные местоимения *(сұраулы-қатыстық есімдіктері)*	*Кім*	ЕС501
	Кімдікі	ЕС502
	Қай	ЕС503
	Қайсы	ЕС504
	Қайда	ЕС505
	Қайдан	ЕС506

Окончание табл. 10

1	2	3
	Қайдағы	EC507
	Қалай	EC508
	Қандай	EC509
	Қанша	EC510
	Қашан	EC511
	Қашаннан	EC512
	Қашанғы	EC513
	Не	EC514
	Немене	EC515
	Неше	EC516
	Нешеу	EC517
	Нешінші	EC518
6. Отрицательные местоимения *(болымсыздық есімдіктері)*	*Еш*	EC601
	Ешкім	EC602
	Ешбір	EC603
	Ештеме	EC604
	Ештеңе	EC605
	Дәнеңе	EC606
	Ешқайда	EC607
	Ешуақытта	EC608
	Ешнәрсе	EC609
	Ешқашан	EC610
	Ешқандай	EC611
	Ешқайсы	EC612
7. Неопределенные местоимения *(белгісіздік есімдіктер)*	*Бір*	EC702
	Біреу	EC703
	Бірнеше	EC704
	Бірдеме	EC705
	Кей	EC706
	Кейбір	EC707
	Кейбіреу	EC708
	Қайсыбір	EC709
	Әр	EC710
	Әрбір	EC711
	Әркім	EC712
	Әрне	EC713
	Әрқайсы	EC714
	Әрқалай	EC715
	Әлде	EC716
	Әлдебір	EC717
	Әлдекім	EC718
	Әлдене	EC719
	Әлдеқайда	EC720
	Әлдеқайдан	EC721
	Әлденеше	EC722
	Әлдеқалай	EC723
	Әлдеқашан	EC724
	Пәлен	EC725

Условно принятые нами коды конкретных местоимений и семантических подклассов местоимений в *табл. 10* записаны в последней графе. EC – сокращение от слова *есімдік*.

Глава четвертая
АБЗАЦ КАК ОСНОВНОЙ СЕМАНТИКО-СИНТАКСИЧЕСКИЙ КОМПОНЕНТ КАЗАХСКОГО ТЕКСТА

4.1. Роль и специфика абзацев в общей структуре текста

Элементарные составляющие микроситуации и слова-наполнители. Известно, что проблема установления смысловой организации является, в целом, семантико-синтаксической задачей [152, с.137]. В то же время, синтаксические особенности этой единицы плана содержания, как отмечают исследователи, нельзя рассматривать в отрыве от ее лексической структуры. Повторяемость и иерархия опорных слов (словосочетаний) в тексте – характерное явление как для общей темы, так и для микротем [4, с.14-15; 57, с.201; 223, с.285; 307, с.107].

В первой главе (пункт 1.5) мы уже выясняли природу СФЕ и абзаца, рассматривая их общность и отличительные черты в плане семантического содержания, а также преимущества и недостатки в определении их границ. Было отмечено, что в абзацах отражаются те связи и отношения некоторого фрагмента психической ситуации, которые зафиксированы у главных и второстепенных референтов с другими, менее значимыми элементами ситуации. К ним можно отнести различные типы действий и состояний главных и второстепенных слов (ГОС и ВОС), различные места действий, время действий, цели, причины действий и состояний и т.п. [208, с.53-54; 307, с.108]. Таким образом, можно выделить два именных ряда соотнесенности: ряд смыслового субъекта и ряд смыслового предиката [326, с.265]. Исследователями предложены различные способы представления иерархии имен, описывающих в пределах микротемы абзаца отмеченные многоаспектные отношения. Одним из этих подходов является дифференцирование слов текста по той роли, которую выполняют их денотаты в определенных ситуациях действительности. Например, такие роли слов, как "часть", "признак", "целое", "принимающий", "действующий", "множество", "средство", "цель" и др. [234, с.164-178; 257, с.112;]. Такой подход обычно связан с фиксацией энциклопедических знаний, которыми располагает автор текста. Предлагается проблему представления энциклопедических знаний рассматривать в неразрывном единстве с проблемой организации словаря текста [257, с.124; 180, с.99]. Из всех подходов, касающихся проблем представления знаний в системах [86, 158, 197], работающих с естественным языком, мы будем придерживаться "ролевого" подхода.

В каждой микроситуации, предположим, можно выделить элементарные составляющие типа "действие", "состояние", "место", "время", "средство", "причина", "цель", "характеризация", "детализация". Этим элементарным составляющим микроситуации (ЭСМС) в тексте абзаца будут соответствовать слова или словосочетания (глаголы, причастия, существительные и их сочетания). Таким образом можно выявить специфику каждой микроситуации, а именно она будет

задаваться в виде определенной последовательности ЭСМС и индивидуальной особенностью отдельного абзаца – наполнением соответствующими денотатами. Итак, в рассматриваемых текстах можно выявить возможные типы абзацев и каждому абзацу приписать в определенном порядке группы слов-конкретизаторов, замещающих в абзаце соответствующие денотаты микроситуаций.

Кроме ГОС, ВОС, ЭСМС и слов-конкретизаторов, в любом тексте можно обнаружить еще одну группу слов, которые "наполняют" полученное содержание качественными признаками, т.е. указывают качества предметов, названных именами существительными, а также признаки действий, определяемые в тексте глаголами и причастиями. Эти слова назовем словами-наполнителями.

Слова, отражающие ЭСМС, в определенной степени, достраивают "костяк" содержания. Такие слова можно задать специальной таблицей основного содержания. Качественная детализация осуществляется в результате предварительной фиксации наиболее вероятных для текстов определенного типа свободных словосочетаний (атрибутивных, адвербиальных, именных). В связи с этим исследователи пользуются такими понятиями, как языковая вероятность, нормирующая вероятность, речевая вероятность и речевая частота [7, с.57]. Если учесть, что сам автор текста, воспринимая внешнюю ситуацию, передает ее в тексте также с некоторыми вероятностями, то возникает естественный вопрос: какие же вероятности свойственны тексту? Для ответа на него необходимо обращение к основам так называемого вероятностного прогнозирования.

Не вдаваясь во все подробности общефизиологических закономерностей, которые изложены в работах [159, с.102-104; 293, с.8], заметим, что такие закономерности справедливы и для речевой деятельности.

Способность человека владеть определенными субъективными представлениями о вероятностях употребления тех или иных элементов речи дает возможность ему строить субъективную модель вероятностной структуры предстоящей ситуации, осуществлять на основе этой модели субъективный прогноз и соответственно строить свою речевую деятельность [154, с.50; 233, с.9-22]. Но, как уже отмечалось, психика человека социальна по своей сути, поэтому речевая способность каждого индивида также социальна. В этой связи, по отношению текста, Г. О. Винокур отмечает, что в каждом тексте может быть выделено нечто "внеличное, общее, надиндивидуальное" [62, с.236].

4.2. Сверхфразовая стереотипичность и абзац

Следует отметить, что использование слов, словосочетаний, предложений и более крупных единиц в процессе порождения текста происходит в ограниченном варианте, а не в бесконечном объеме. Такое ограничение объема лингвистических единиц практически есть и проявление социального характера речевой способности человека. При этом, для каждого индивида такое ограничение объема единиц может быть разным. Это, видимо, объясняется воздействием на человека очень многих внешних и внутренних факторов, действующих на личность на протяжении всей его жизни [87, с.83; 279, с.202; 338, с.33]. У человека основа такой совокупности единиц (ядро) формируется еще с детства под влиянием окружающей среды – коллектива, например, в процессе подражания взрослым [45, 247, 354]. В тексте же такое "ядро" создается в явлении стеоротипии, проявляющееся на разных

уровнях [48, 66, 243, 286, 340]. Как утверждают ученые, стереотипичность, или стандартность речевого действия, играет в основном роль экономии усилия как автора, так и воспринимающего речь индивида [313, с.135; 340, с.223]. Стереотип усреднен исторически из индивидуальных решений и находится в рамках нормативной грамматики [286, с.252; 313, с.315; 340, с.223].

Такие стереотипы, как фонематический, словесный, фразовый и текстовый [237, с.90-98], служат как бы пусковым звеном в развертывании соответствующего фрагмента речевой программы. Необходимо отметить, что с повышением рассматриваемого уровня, языковой стереотип становится менее стабильным, получая большее разнообразие форм воплощения [348].

Рассмотрим более подробно проблему фразового стереотипа. Известно, что каждому отраженному в сознании фрагменту действительности в некоторое соответствие ставятся определенные языковые формы. Такое соответствие, в основном, является результатом неразрывности мышления и языка [153, с.44; 305, с.43]. Для одних исследователей роль таких языковых форм играют фиксированные слова [61, с.64], для других – словосочетания, предложения и даже целые тексты [90, с.16]. Многие специалисты (психологии, психолингвистики, логики и лингвистики) отмечают, что в процессе человеческого познания наиболее важные и часто повторяющиеся отношения реальной действительности закрепляются в форме синтаксических структур. Такой процесс усвоения знаний начинается у человека с детства вместе со словами и звуками родного языка. Например, в процессе "говорения" очень часто человек, не задумываясь, выбирает один из наиболее привычных для него в данной ситуации языковых шаблонов [6; 114, с.145].

Такие языковые шаблоны в научной литературе встречаются в различной терминологии: "синтаксические модели", "структурные схемы предложений", "конструкции", "фразовые стереотипы", "синтаксические формы" и т.д., и т.п. Как бы они не назывались, все они являются психическими реальностями и участвуют в процессе создания текста вместе с единицами других уровней языка.

Обычно внеязыковое содержание чаще всего передается в тексте (устном или письменном) не одним, а несколькими взаимосвязанными предложениями. Такое утверждение экспериментально подтверждается и психологами, и лингвистами. Таким образом, в языке существуют определенные структурно-смысловые модели для выражения отношений между моделями в структурах, находящихся за пределами предложения [114, с.146; 183, с.61].

Если говорить о более углубленных исследованиях в этой области, то наличие таких синтаксических шаблонов связано с общим положением об опережающем отражении действительности, разработанным советскими учеными Д. Н. Узнадзе и П. К. Анохиным [10]. Согласно этой теории, человек "обладает способностью, на основе врожденного и приобретенного опыта, предвидеть развитие ситуативных обстоятельств и настраиваться на определенную реакцию с некоторым упреждением вероятностного хода событий" [154, с.50].

Такие синтаксические единицы в современной теории текста – в лингвистике текста называют "сверхфразовыми единицами" или "сложными синтаксическими целыми". Многими исследованиями последних лет допускается, что и абзац письменного текста является "логическим слепком" с действительности [4, 145, 183, 225, 254, 268, 286, 294]. Такой подход к природе абзаца подтверждается не только лингвистическими исследованиями, но и экспериментами психологов и психолингвистов [51, 269, 342]. Приведенные факты в этих работах показывают, что абзацы выделяются автором не произвольно, а по определенным правилам.

Для научных текстов эти правила носят строгий и стереотипный характер, тогда как для художественных текстов – более свободный и творческий. "В то же время, – замечает И. Я. Чернухина, – в особенностях композиции абзаца отражается семантическая динамика текста" [306, с.155]. Наряду с этим, во многих исследованиях текста на уровне абзаца отмечаются специфичность абзацев для каждого автора произведения [145, с.258-259; 255, с.210].

Таким образом, стереотипичные фразовые построения в нашем изложении будут называться "шаблонами" или "синтаксическими моделями".

С точки зрения речевой деятельности, фразовая стереотипичность это не что иное, как "вызов из памяти некоторой общей схемы развития мысли, а следовательно, и его речевого оформления" [154, с.46]. Под влиянием стереотипичности на уровне фраз в каждой фразе текста актуализация одних лексических элементов неизбежно сопровождается автоматизацией выбора других элементов [67, с.60; 345, с.73].

Роль и специфика абзацев в общей структуре текста, о которых было упомянуто, дает нам также основание полагать и о сверхфразовой стереотипичности. В повседневной жизнедеятельности, т.е. в процессе обучения жизни и языку, несомненно, в нашем сознании откладываются основные характеристики различных текстов, представляющих научный, публицистический, художественный, деловой и другой жанр. Эти характеристики в совокупности представляют собой не что иное, как вероятностные модели и представления о синтаксических структурах указанных текстов. В результате влияния стереотипичности на фразовом и сверхфразовых уровнях индивидуальные различия внутри одного вида письменной речи становятся менее значимыми, чем, скажем, различия между разными жанрами письменных текстов, созданных одним и тем же автором [181, с.76; 219, с.113]. Видимо, такая сверхфразовая стереотипичность, в какой-то степени, связана с тем положением, что порождение текста и его восприятие происходят определенными "блоками", "квантами", в которых сохранена связь с предшествующим и заложено на некоторое семантическое расстояние последующее [67, с.60; 50; 154, с.50; 206, с.82].

В силу того, что в самой основе речевой деятельности находится прежде всего вещественное содержание речи, то, нужно полагать, сама коммуникация дискретна и таким образом деление текста на определенные "кванты" или "куски" основано на "логическом слепке" с действительности в результате обобщения человеческого опыта научного познания [79, с.3; 270, с.111; 307, с.107; 314, с.72]. Во многих работах показано, что сложные формы умственной деятельности также слагаются из элементарных информационных процессов [47, с.481; 124, с.15; 220, с.83].

Объединение в единое целое указанных выше "кусков" информации возможно по следующим трем принципам: 1) вероятностном, если объединение происходит на основе вероятностных распределений; 2) детерминированном, когда порядок следования "кусков" однозначен и строго задан; 3) вероятностно-алгоритмическом, когда в процессе объединения учитываются и случайные, и детерминированные (алгоритмические) связи.

Последний случай объединения, т.е. вероятностно-алгоритмический, выступает как основной тип процессов объективного мира [244, с.45].

"Квантование" смысла текста во многом зависит от типа текста, т.е. от его функциональной разновидности, направленности, намерений автора, степени его языковых знаний, т.е. от профессиональных качеств знаний языка и ряда других факторов. В целях изучения семантической структуры текстов, многими исследователями приходилось выделять различные смысловые "куски". Авторы

этих работ именовали их как "текстовые элементы", "блоки-мотивы", "фрагменты", "аспекты", "сгустки" и "абзацы" [78, с.7; 209, с.200; 280, с.132; 319, с.209].

Таким образом, абзац, как смысловой "квант" и как "застывший" синтаксический шаблон, может быть принят за минимальный логико-семантический компонент текста.

Многоплановость содержания текста может быть понята до конца только тогда, когда мы детально рассмотрим текст и его основные единицы с точки зрения семантики.

В последующем, опираясь на описанные особенности текстовых единиц, сделана попытка формального представления структуры и содержания казахского текста.

4.3. Функционально-смысловая классификация казахского текста

В предыдущих разделах мы старались показать специфику абзаца как некоторого смыслового кванта общего содержания текста и как некоторого синтаксического шаблона, отражающего отношения определенного фрагмента реальной ситуации.

Для построения модели структуры казахского текста мы должны снабдить ее смысловыми квантами и синтаксическими шаблонами, показать их специфику для различных функциональных стилей.

Далее наши исследования будут посвящены поиску и выделению таких специфичных особенностей абзацев, которые помогли бы нам построить алгоритм, объединить их в единый связный текст.

Основная особенность текста любого языка, в том числе, и казахского, заключается в том, что между предложениями (и их комбинациями) устанавливаются смысловые связи на основе отношений между объектами и явлениями действительности. При этом одни предметы или явления действительности объединяет их последовательность во времени, другие – пространственно или иным образом, т.е. один предмет (или явление) служит причиной другого, например третий выступает следствием первого [198, с.116].

Ученые отмечают, что в тексте предложения интегрируются в типы высказывания или типы речи (описание, повествование, рассуждение), каждый из которых имеет определенное функционально-смысловое значение и стабильную структурно-языковую характеристику [204, с.228].

Приведем определения для каждого типа речи в отдельности, т.е. для повествования, описания и рассуждения:

1. Если фразы или более крупные текстовые единицы расположены в тексте в порядке, определяемом связью во времени описываемых в тексте событий, то такой тип связи называют повествованием *(баяндау)*.

Можно говорить о таких структурных разновидностях повествования, как: а) конкретно-сценическом; б) обобщенно-сценическом; в) информационном. Повествование можно назвать конкретно-сценическим *(нақты-сценалық)*, если в тексте речь идет о расчлененных, хронологически последовательных действиях (или состояниях) одного или нескольких субъектов в данной конкретной обстановке в пределах одной микротемы.

Обобщенно-сценическое повествование *(жалпылама-сценалық)* это такой вид повествования, которое сообщает о конкретных последовательных действиях, причем эти действия повторяются в данной обстановке и становятся типичными для нее.

Последовательность действий при этом передается семантикой глаголов, их видовременными формами, наличием определенных обстоятельственных слов и другими средствами.

Для информационного повествования характерно наличие в тексте сообщения о каких-либо действиях (или состояниях) без их конкретизации и детальной хронологической последовательности.

2. Если связи между фразами или более крупными единицами текста осуществляются на основе пространственных, функциональных или иных отношений между частями какого-либо отражаемого в речи объекта, то такой тип речи называется описанием *(суреттеу)*. Описание – это констатация наличия или отсутствия каких-либо признаков или явлений.

Основное отличие описания с грамматической точки зрения состоит в том, что в нем в основном используются только изъявительное наклонение *(ашық рай)*, ограниченное число вводных слов, отмечается особое употребление предикативно-модальных слов. Последние указывают, как правило, на отношение между действием и субъектом.

Важным для нас представляются те структурно-смысловые разновидности, которые выделяются в описании [204, с.60-70]:

1) различные описания природы;

2) описание внутренней части какого-либо помещения;

3) описание внешнего вида какого-либо одушевленного существа или неодушевленного предмета (описание портрета человека и его одежды, описание животного, растения, зданий, предметов и т.д.);

4) описание-характеристика какого-либо одушевленного существа или неодушевленного предмета (например, черты характера человека, рода его занятий, его наклонности и т.п., при описании-характеристике предмета подчеркиваются некоторые свойства этого предмета).

3. Если фразы или более крупные единицы текста связаны причинно-следственными выводными отношениями, то такой тип речи называется рассуждением *(пайымдау)*.

Рассуждение как функционально-смысловой тип речи в отличие от рассмотренных, в логическом плане выражает собой некоторое умозаключение, т.е. высшую форму логического мышления. Можно утверждать, что рассуждение, относясь к логической категории, является и определенной структурно-смысловой единицей текста [204, с.150].

В результате анализа научно-популярных казахских текстов нам удалось дифференцировать несколько видов рассуждений и выделить те языковые формы, которые для них им характерны. По нашим статистическим данным, наиболее употребительны следующие шесть видов рассуждений:

1) оценочные именные рассуждения;

2) рассуждения со значением состояния;

3) рассуждения с обоснованием реальных или гипотетических действий;

4) рассуждения со значением необходимости, возможности или желательности некоторого действия;

5) рассуждения с заранее обусловленными действиями;

6) рассуждения с категорическим отрицанием или утверждением, они выражаются в тексте в форме риторического вопроса или восклицания.

Представленные особенности повествования, описания и рассуждения наиболее ярко проявляются в пределах наименьших семантико-синтаксических единиц, т.е.

абзацев. Поэтому в последующем анализе будем выделять такие разновидности абзацев, как абзацы повествования, абзацы описания и абзацы рассуждения [204, с. 232].

Практически в текстах любого типа может встречаться диалог. При этом каждую смену реплики без косвенной речи автора и косвенную речь, сопровождающую реплику (вместе с репликой), мы решили считать за абзац. Поэтому, в дальнейшем, примем, что существуют абзацы с вопросительной речью и абзацы с побудительной речью. Кроме того, возможны случаи, когда в одном и том же абзаце присутствуют признаки двух или более перечисленных функциональных типов (описание, повествование, рассуждение). Такие абзацы мы решили называть абзацами смешанного типа [204, с.229].

Итак, в дальнейшем будем различать следующие типы абзацев:

1) повествовательные *(баяндау);*

2) описательные *(суреттеу);*

3) типа рассуждения *(пайымдау);*

4) абзацы с выражением вопроса *(сұрау мәнді);*

5) абзацы с выражением побуждения *(ой тастау, әсер ету мәнді);*

6) абзацы смешанного типа *(аралас типті).*

Для некоторых видов текстов исследователи обнаруживают специфику в структуре абзацев, начинающих текст, находящихся в его середине и заканчивающих текст.

Как известно, начальные абзацы текста являются некоторым прологом к последующему изложению, излагая точку зрения автора на ту или иную проблему. Конечные или заключительные абзацы, как правило, подводят итог изложенному или служат своеобразным послесловием, позволяющим до конца уяснить мысль автора [205]. Поэтому, с опорой на такую классификацию, мы выделяли начальные, медиальные и конечные абзацы во всех трех нами рассматриваемых казахских текстах.

В результате статистических подсчетов выяснилось, что каждый функционально-смысловой тип, в большей или меньшей степени, может встретиться в тексте как в отдельности, так и в смешанном виде. Эти результаты приводятся нами в *табл. 11 – 13.*

В кратце проанализируем данные указанных таблиц.

1. По художественному типу изучено 20 текстов, состоящих из 487 абзацев. Согласно статистическим данным (*табл. 11*), для художественных текстов в начальных абзацах характерным является тип повествования, так как из 44 начальных абзацев 30 (68%) относятся к данному типу. Рассматривая далее, убеждаемся в том, что в начале текста только 9 абзацев, т.е. 20,4% из всего числа начальных абзацев, относятся к смешанному типу абзацев (в абзаце участвуют какие-нибудь два типа из трех или все три типа одновременно). По данным *табл. 11,* остальные функционально-смысловые типы для художественго текста в начальных абзацах в рассматриваемых нами случаях не характерны. Тип повествования свойствен и медиальным, и заключительным абзацам художественого текста – 54 и 69% соответственно. В медиальных 404 абзацах можно встретить, в той или иной степени, все 6 видов перечисленных абзацев. Например, наибольший процент встречаемости после типа повествования занимает смешанный вид абзацев и далее, в порядке убывания, следуют абзацы со значением побуждения, абзацы со значением вопроса, рассуждения и описания. Для 39 конечных типов абзацев в художественном тексте картина иная, хотя первенствует, так же, как и в приведенных случаях, тип повествования (69%). На втором месте по частоте встречаемости

находятся абзацы со значением побуждения (15,4%), на третьем – рассуждения (10,3%), далее следуют смешанные по типу абзацы, а абзацы со значением вопроса вообще не отмечены.

2. В *табл. 12* показана статистика структурных типов абзацев по газетным текстам казахского языка. По данным *табл. 12* можно отметить, что функционально-смысловой тип – повествование (так же, как и в текстах художественного типа) имеет преимущество в частотном отношении по сравнению с другими функционально-смысловыми типами абзацев, а именно в начальных и в конечных абзацах в 47 публицичтических текстах наблюдался в основном тип повествования – 89 и 96% соответственно. Однако, в средней части текста, т.е. в медиальных абзацах, можно встретить все виды функционально-смысловых типов, опять таки, с характерным перевесом типа повествования. На втором месте после повествования стоит смешанный вид абзацев (17%), т.е. абзацы, которые можно отнести к типу как повествования, так и описания или же к типу рассуждения. Тип абзацев, относящихся к абзацам-рассуждения, занимает здесь только третье место. Далее, в порядке убывания частоты, следуют: абзацы-описания (9%), абзацы-вопросы (2%) и абзацы-побуждения (1%).

В 35 газетных текстах абзацы имеют заключающий характер, т.е. из рассматриваемых 57 конечных абзацев 55 относятся к типу повествования – это 96% всех конечных абзацев, а оставшиеся 4% распределены поровну между типами рассуждения и смешанным видом – по одному абзацу на каждый тип.

3. Рассмотрим статистику функционально-смысловых типов абзацев в 47 научно-популярных казахских текстах *(табл. 13)*. Сравнив результаты подсчета по этому типу текстов с изложенными двумя типами (т.е. с художественным и публицистическим), можно отметить, что и в этом случае общая закономерность, в основном не меняется. Функционально-смысловой тип абзацев как повествование по частоте встречаемости имеет преимущество во всех рассматриваемых структурных типах (т.е. начальных, медиальных и конечных абзацах). Округляя данные, отметим, что 70% начальных, конечных и около 50% всех медиальных абзацев относится к типу повествования. Далее, среди начальных абзацев второе место принадлежит смешанному виду абзацев (17%), т.е. это, в основном, тот случай, когда абзацы по содержанию можно отнести и к типу повествования, и к типу рассуждения (сокращенно П+Р). Следует подчеркнуть, что среди смешанных видов абзацев случай П+Р является более характерным, как для медиальных абзацев, так и для конечных абзацев в рассматриваемых научно-популярных текстах. В медиальных абзацах научно-популярных текстов после повествования на второе место по частоте встречаемости можно поставить тип рассуждения (28.4%), тогда как среди конечых абзацев на это место претендует смешанный вид абзацев – "повествование+ рассуждение".

В крайних позициях текстов, т.е. в начальных и конечных абзацах, тип рассуждение, уступая другим типам, занимает уже третье место по частоте встречаемости в тексте (5 и 13% соответственно).

В заключение хотелось бы отметить, что все изложенное по поводу занимаемых мест по частоте встречаемости того или другого функционального-смыслового типа в структурном отношении (т.е. в позиционном расположении абзацев) не претендует на стабильность, так как это может зависеть от многих факторов: от характера текстов, индивидуальных особенностей исследователя (т.е. от уровня знания той области, о которой идет речь в тексте), который должен отнести тот или иной абзац к одному из функционально-смысловых типов, а также от возможности считать его или начальным, или медиальным, или конечным абзацем.

Таблица 11. **Функционально-смысловые типы абзацев в художественных текстах казахского языка**

№ п/п	Функционально-смысловой тип абзаца	Структурный тип абзаца и частота их употребления					
		начальный		медиальный		конечный	
		абсолют. F	относит. f	абсолют. F	относит. f	абсолют. F	относит. f
1	Повествование (П)	30	0,682	220	0,544	27	0,692
2	Описание (О)	1	0,023	9	0,022	-	-
3	Рассуждение (Р)	3	0,068	17	0,042	4	0,103
4	Смешанные виды: П+О;П+Р;О+Р;П+О+Р	9	0,204	84	0,208	2	0,051
5	Абзацы со значением вопроса (В)	1	0,023	28	0,069	–	–
6	Абзацы со значением побуждения (Б)	–	–	46	0,114	6	0,154
	Итого	44	1,000	404	1,000	39	1,000

Таблица 12. **Функционально-смысловые типы абзацев в публицистических текстах казахского языка**

№ п/п	Функционально-смысловой тип абзаца	Структурный тип абзаца и частота их употребления					
		начальный		медиальный		конечный	
		абсолют. F	относит. f	абсолют. F	относит. f	абсолют. F	относит. f
1	Повествование (П)	40	0,890	234	0,580	55	0,960
2	Описание (О)	1	0,020	36	0,090	-	-
3	Рассуждение (Р)	-	-	51	0,130	1	0,020
4	Смешанные виды П+О;П+Р;О+Р;П+О+Р	4	0,090	70	0,170	1	0,020
5	Абзацы со значением вопроса (В)	–	–	9	0,020	–	–
6	Абзацы со значением побуждения (Б)	–	–	3	0,010	–	–
	Итого	45	1,000	403	1,000	57	1,000

Таблица 13. **Функционально-смысловые типы абзацев в научно-популярных текстах казахского языка**

№ п/п	Функционально-смысловой тип абзаца	Структурный тип абзаца и частота их употребления					
		начальный		медиальный		конечный	
		абсолют. F	относит. f	абсолют. F	относит. f	абсолют. F	относит. f
1	2	3	4	5	6	7	8
1	Повествование (П)	71	0,763	199	0,488	53	0,679
2	Описание (О)	1	0,011	36	0,088	2	0,026
3	Рассуждение (Р)	5	0,054	116	0,284	10	0,128
4	Смешанные виды	16	0,172	57	0,140	13	0,167
	Из них: П+О	4	0,043	3	0,007	2	0,026
	П+Р	11	0,118	41	0,100	10	0,128
	О+Р	1	0,011	10	0,024	1	0,013
	П+О+Р	–	–	3	0,007	–	–
5	Абзацы со значением вопроса (В)	–	–	–	–	–	–
6	Абзацы со значением побуждения (Б)	–	–	–	–	–	–
	Итого	93	1,000	408	1,000	78	1,000

4.4. Классификация абзацев казахского текста по их предметно-логическому содержанию

О предметно-логическом содержании абзаца.

Как мы уже отмечали, абзац содержит описание некоторой микроситуации, т.е. определенного числа объектов реального мира с их свойствами в их взаимоотношениях. Выделение групп предложений в абзацы не является произвольной операцией автора текста, а определяется рядом факторов: необходимостью изложения новой мысли, важностью передаваемой информации в рамках всего текста, невозможностью дальнейшего линейного представления новых сведений, сменой речевых субъектов и т.д. [183, с.86-87].

Сушествуют различие подходы к оценке количественной стороны микротем, участвующих в плане содержания абзаца. Одни ученые говорят о монотематиченности, а другие, наоборот, о политематиченности [252, с.93, 308, с.5-6]. Анализ нашего материала также показывает, что абзацы реальных казахских текстов вполне могут быть политематичными.

Рассмотренные функционально-смысловые типы абзацев представляют собой ни что иное, как наиболее общие типовые способы изложения содержания. Дальнейшей конкретизацией содержания абзаца является понятие предметно-логического содержания абзаца [83, 172].

В соответствии с принятым нами подходом к строению текста, предполагающему наличие статической и динамической составляющей в тексте, для нас важно, какие именно факты объективной действительности перечисляются в абзаце-описании, абзаце-рассуждении и абзаце-повествовании словами: выраженные главными опорными словами (ГОС), второстепенными опорными словами (ВОС) или прочими словами (ПС) текста. Поэтому, уточняя имеющие в литературе определения "предметно-логического содержания абзаца" для отдельных функционально-смысловых типов, примем их в следующей формулировке:

1. Предметно-логическим содержанием абзаца-описания называется дифференцированное по типам главных и второстепенных субъектов, объектов, мест и времен микроситуации, описанной в абзаце, перечисление фактов объективной действительности (признаков, явлений, событий, состояний взглядов, теорий и т.п.), связанных с содержанием всего текста.

2. Предметно-логическим содержанием абзаца-рассуждения называется дифференцирование по типам главных и второстепенных субъектов, объектов, мест и времени ситуации, представленной в тексте, последовательная связь мыслей, связанных с содержанием всего текста.

3. Предметно-логическим содержанием абзаца—повествования можно назвать дифференцированное по типам главных и второстепенных субъектов, объектов, мест и времен ситуации, представленной в тексте, последовательное изложение в абзаце действий, событий, состояний с детализацией или без нее.

Как показывает наш практический анализ выделения абзацев с определенным предметно-логическим содержанием, это – весьма трудная и неоднозначная задача. Один и тот же абзац может быть отнесен и к функционально-смысловому типу описания или повествования, а также к рассуждению или описанию и т.д. Несмотря на указанные трудности, мы согласно сформулированным определениям предметно-

логического содержания абзацев, для каждого из рассматриваемых стилей провели классификацию типов абзацев по их предметно-логическому содержанию. При позиционной классификации (т.е., рассматривающей абзацы в виде начальных, медиальных и конечных) нами были исследованы тексты только двух стилей: научно-популярного и публицистического, а в художественном стиле классификация абзацев на предметно-логическое содержание проводилась без учета их позиции в тексте. Результаты статистического и классификационного исследования отражены в виде семи таблиц: 1) *табл. 14, 15, 16* – для абзацев в научно-популярных текстов; 2) *табл. 17, 18, 19* – для абзацев публицистических текстов; 3) *табл. 20* – для абзацев художественных текстов. Рассмотрим в кратце статистические данные указанных таблиц.

4.5. Особенности абзацев научно-популярного текста по их предметно-логическому содержанию

Исследователи отмечают, что научный текст на всех его уровнях членится с опорой на "логический слепок" с действительности на основе обобщения человеческого опыта научного познания [254, с.27-35]. Ввиду общности структуры научного исследования, фиксированности этапов научного познания и преобладания логических способов в научном рассуждении выделение абзацев в научном тексте происходит почти всегда одинаково (стереоптично) [239].

Научно-популярные тексты, которые нами взяты для эксприментальных исследований, таким единством организации (стереотипичностью) не отличаются. В то же время, в их основе лежит определение "однородность" текста в тематико-ситуационном, пространственном и временном плане [271].

Рассматривая типы начальных абзацев, можно отметить, что в результате классификационного анализа 47 казахских научно-популярных текстов нами было выявлено 93 начальных абзацев, которые составили 20 типов по их предметно-логическому содержанию *(см. табл. 14)*.

Таблица 14. **Типы начальных абзацев научно-популярного текста по их предметно-логическому содержанию**

№ п/п	Типы начальных абзацев	Код типов	Частота	
			F	f
1	2	3	4	5
1	Констатация некоторого состояния главного объекта в зависимости от других главных и второстепенных объектов	М001	6	0,0645
2	Утверждение о состоянии главного объекта и констатация будущих действий автора	М002	1	0,0108
3	Констатация некоторой специфичности главного объекта в связи с другими главными и второстепенными объектами	М003	11	0,1183
4	Подчеркивание некоторой специфичности главного объекта и краткая констатация действий главного субъекта	М004	13	0,1398
5	Констатация некоторого прошлого действия автора	М005	6	0,0645

Окончание табл. 14

1	2	3	4	5
6	Утверждение о сути некоторого главного объекта	M006	10	0,1075
7	Высказывание некоторого условия, связанного с будущим действием автора	M007	10	0,1075
8	Подчеркивание некоторой специфичности главного объекта и ее влияние на главный субъект и другой главный объект	M008	5	0,0538
9	Констатация некоторого действия автора, причины этого действия и оценка будущих действий автора	M009	5	0,0538
10	Высказывание предположения об особенностях главного субъекта и условиях изменения этих особенностей	M010	1	0,0108
11	Констатация некоторого простого утверждения автора, связанных с главным объектом	M011	9	0,0968
12	Описание некоторых особенностей главного объекта и действий прочих субъектов, связанных с этим главным объектом	M012	5	0,0538
13	Утверждение об особенностях прочих субъектов, связанных с главным объектом	M013	1	0,0108
14	Констатация некоторого последующего действия автора, связанного с характеристикой главного объекта	M014	3	0,0323
15	Высказывание предположения о сути главного объекта	M015	2	0,0215
16	Оценка прошлого действия автора, констатация его будущего действия и постановка вопросов, связанных с главными и второстепенными объектами	M016	1	0,0107
17	Подчеркивание особенностей второстепенных объектов и описание действий автора	M017	1	0,0107
18	Двустороннее описание особенностей главного объекта	M018	1	0,0107
19	Констатация последующих действий автора, связанных с главным объектом	M019	1	0,0107
20	Констатация прошлого действия автора, связанного с характеристикой главного объекта и констатация предстоящего действия автора	M020	1	0,0107
		ВСЕГО	93	1,0000

Как видно из *табл. 14*, наиболее употребительными из типовых начальных абзацев являются следующие:

– "подчеркивание некоторой специфичности главного объекта и краткая констатация действий главного субъекта" (M004; f=0,1398);

– "констатация некоторой специфичности главного объекта в связи с другими главными и второстепенными объектами" (M003; f=0,1183);

– "высказывание некоторого условия, связанного с будущим деиствием автора" (M007; f=0,1075).

По существу, это наиболее типичное содержание начальной части научно-популярного текста, представляющего главный объект исследования и то, что сделано или будет сделано автором для раскрытия сути этого объекта.

В качестве примера рассмотрим один начальный абзац научно-популярного текста по теме ***"Кенді ауыр суда байыту":***

"Тығыздығы 1 кг/м³ ден асатын кез келген зат суда батады. Мұндай заттардың өте ұсақ түйіршіктерінің өзі суда қалқуы мүмкін емес. Ал мына көріністің мәні ерекше: жұдырықтай кесектер суда қалытқыдай қалқуда... Мұның мәні – жай судың орнына салмағы ауыр су алынуында...".

В данном абзаце подчеркивается некоторая специфичность главного объекта, т.е., то, что вещества с плотностью более 1 кг/м³ тонут в воде. Далее идет краткая констатация действий главного субъекта (ученого, экспериментатора), который должен действовать так, чтобы вещество не утонуло в воде, т.е. увеличить плотность воды более, чем на 1 кг/м³.

Таким образом, приведенный абзац относится к типу: "подчеркивание некоторой специфичности главного объекта и краткая констатация действий главного субъекта". Подобный же анализ 353 медиальных абзацев научно-популярного текста позволил нам выделить 23 типовые разновидности по предметно-логическому содежанию абзацев *(табл. 15)*.

Как видно из *табл. 15*, в экспериментальных научно-популярных текстах более употребительными среди типов медиальных абзацев по предметно-логическому содержанию оказались следующие:

– "подчеркивание специфичности второстепенного объекта"(М111, f = 0,1445);

– "констатация некоторого прошлого и настоящего действия автора" (М113, f = 0,1048);

– "констатация некоторых условий, связанных с прочими второстепенными объектами" (М110, f = 0,1048);

– "констатация прошлых действий автора и описание поведения главного субъекта" (М123, f = 0,0652);

– "констатация состояния главного объекта" (М122, f=0,0537);

– "краткая двусторонняя характеристика поведения главного субъекта" (М118, f = 0,0482).

Приведенное предметно-логическое содержание абзацев свидетельствует о том, что в средней, т.е. в основной части текста, происходит сам процесс изучения главного объекта путем анализа действий (поведения) главных субъектов и некоторых действий (физических и умственных) автора, проводящего исследование.

В качестве примера медиального абзаца научно-популярного текста рассмотрим абзац *(Н10-08)*: *"Магнитті байытудың өнімділігін арттырудың бір жолы - магниттілігі аз минералдардың магниттік қасиетін алдын ала күшейту. Мысалы, гематит, сидерит және лимонит секілді минералдардың магниттілігі өлсіз. Егер осы минерал рудаларын 550-600 градуста көміртегін қосып күйдірсе күшті магниттеледі".*

По содержанию данный абзац, вероятно, можно отнести к типу: "констатация некоторых условий, связанных с прочими второстепенными объектами".

Таблица 15. **Типы медиальных абзацев научно-популярного текста по их предметно-логическому содержанию**

№ п/п	Типы медиальных абзацев	Код типов	Частота	
			F	f
1	2	3	4	5
1	Констатация некоторых условий, определяющих будущие действия автора	М101	4	0,0113
2	Констатация действия автора в связи с действиями главного субъекта и главного объекта	М102	5	0,0142
3	Описание некоторых прошлых действий автора	М103	1	0,0028
4	Описание (обоснование) причины особой организаций главно-го объекта в связи с другим главным объектом и главным субъектом	М104	5	0,0142
5	Утверждение автора о сути прочего субъекта	М105	11	0,0312
6	Констатация причины специфичности главного объекта и постановка вопросов, связанных с главным субъектом	М106	3	0,0085
7	Описание состава главных объектов и констатация причины именно этого состава	М107	19	0,0538
8	Описание состава главных субъектов	М108	6	0,0170
9	Некоторая характеристика действий второстепенного объекта	М109	13	0,0368
10	Констатация некоторых условий, связанных с прочими второстепенными объектами	М110	37	0,1048
11	Подчеркивание специфичности второстепенного объекта	М111	51	0,1445
12	Утверждение автора о некоторых особенностях главного объекта и описание их особенностей	М112	9	0,0255
13	Констатация некоторого прошлого и настоящего действия автора	М113	37	0,1048
14	Постановка вопросов, связанных с главным субъектом	М114	8	0,0227
15	Констатация промежуточного вывода о поведении главного субъекта и описание способа изменения этого поведения	М115	14	0,0397
16	Краткая промежуточная и итоговая характеристика главного субъекта	М116	41	0,1161
17	Подчеркивание некоторой особенности главного субъекта в связи с второстепенным объектом	М117	11	0,0312
18	Краткая двусторонняя характеристика поведения главного субъекта	М118	17	0,0482
19	Утверждение об особенностях главного объекта с выходом на рисунок (таблицу и т.п.) и характеристика последнего	М119	1	0,0028
20	Констатация особенности главного субъекта и ее изменение при изменений действий автора	М120	14	0,0397
21	Констатация отдельного предстоящего действия автора	М121	4	0,0113
22	Констатация состояния главного объекта	М122	19	0,0537
23	Констатация прошлых действий автора и описание поведения главного субъекта	М123	23	0,0652
		Всего	353	1,0000

Рассмотрим другой пример из (Н41-06) (т.е. 6 абзац из текста с номером 41 по научно-популярному типу текстов): *"Микроорганизмдердің тіршілік қызметі үшін қалдық суларда азот, фосфор мен калийдің болуы қажет, сондықтан бұл элементтер болмаған кезде оларды тиісті минерал тұздар түрінде қосып отырады".*

Содержание этого абзаца можно передать так: *"Для жизнеспособности микроорганизмов в водоемах необходимым условием является присутствие в составе воды таких химических элементов, как азот, фосфор и калий. При недостаточности каких-либо из этих элементов их добавляют в виде солей соответствующих минералов".*

Таким образом, учитывая все перечисленные "условия для жизни микроорганизмов", мы решили абзацы подобного содержания отнести к типу "подчеркивание некоторой особенности главного субъекта в связи с второстепенным объектом".

Данные *табл. 16* являются результатом анализа 57 конечных абзацев 47 научно-популярных казахских текстов. Целью исследования является выявление особенностей конечных абзацев по их предметно-логическому содержанию. В *табл. 16* указаны 10 таких типовых абзацев, которые имеют кодовые номера от М201 по М210 с указанием на их частоты употребления в тексте. По нашим статистическим данным, в конечных (или заключительных) абзацах к самым употребительным типам в отношении предметно-логического содержания можно отнести следующие:

– "заключительное итоговое утверждение о сути главного объекта" (М201, f =0,3333);

– "некоторая характеристика главного объекта с постановкой вопроса и оценка этого вопроса" (М208, f = 0,2283);

– "утверждение автора об особенностях главных объектов" (М202, f = 0,1754).

Подобное содержание конечных абзацев вполне объяснимо, так как в заключении любого научного текста, как правило, формулируются основные выводы, результаты проведенного исследования и намечаются действия автора (исследователя) по дальнейшему развитию исследования.

Таблица 16. **Типы конечных абзацев научно-популярного текста по их предметно-логическому содержанию**

№ п/п	Типы конечных абзацев	Код типов	Частота	
			F	f
1	Заключительное итоговое утверждение о сути главного объекта	М201	19	0,3333
2	Утверждение автора об особенностях главных объектов	М202	10	0,1754
3	Констатация прошлых действий автора, связанных с главным объектом	М203	3	0,0526
4	Постановка некоторых вопросов и констатация будущих действий автора	М204	2	0,0351
5	Описание будущих действий автора в связи с главным объектом	М205	1	0,0175
6	Заключительное утверждение, связан. с главным субъектом, с выходом на рисунок (график и т.п.)	М206	1	0,0175
7	Констатация предстоящих действий автора	М207	1	0,0175
8	Некоторая характеристика главного объекта с постановкой вопроса и оценка этого вопроса	М208	13	0,2283
9	Краткая двусторонняя характеристика главного объекта	М209	4	0,0702
10	Высказывание предположений о будущем поведении главных и второстепенных объектов	М210	3	0,0526
		Всего	57	1,0000

Приведем пример абзаца казахского научно-популярного текста, содержание которого по функционально-смысловому характеру относится к абзацу-повествованию, а по предметно-логическому содержанию к типу "некоторая характеристика главного объекта с постановкой вопроса и оценка этого вопроса".

Пример: *"Атмосфера ауасының тазалығы жолында күрес жүргізу – ең алдымен адамдардың денсаулығын сақтау мәселесі. Сондықтан, адамдардың денсаулығына бірінші дәрежелі маңыз аударылатын елде – атмосфера ауасының гигиенасы гигиеналық ғылымның арнаулы саласы ретінде құрылып, ойдағыдай дамып келеді".*

Как видно из примера, главным объектом исследования взят "атмосферный воздух" *(атмосфера ауасы)*. В первом предложении абзаца автор старается дать главному объекту некоторую характеристику, т.е. говорится о тесной взаимосвязи чистоты атмосферного воздуха и здоровья людей. Во втором заключительном предложении абзаца ставится вопрос о необходимости научного его изучения и дается оценка о том, что такая наука "о гигиене" есть и она неплохо развивается у нас в стране.

Итак, в организации содержания научного текста, независимо от его структурной части, существуют определенные общие закономерности, выражающиеся в употреблении в разных частях такого текста абзацев с одним и тем же предметно-логическим содержанием.

4.6. Особенности абзацев публицистического текста по их предметно-логическому содержанию

Аналогично изложенному в пункте 4.5, мы выделили всевозможные типы абзацев по их предметно-логическому содержанию в 35 публицистических (газетных) текстах. Как было отмечено, это статьи из газеты "Қазақстан мұғалімі" ("Учитель Казахстана"), посвященные одной тематике: "Алтыжасарлар мектепке барады" (по рубрике "Шестилетние идут в школу").

Результаты классификационного анализа для начальных, медиальных и конечных абзацев приведены соответственно в *табл. 17, 18* и *19*. Рассматривая, по аналогии с ранее рассмотренным научно-популярным текстом, особенности отмеченных газетных текстов по употребительности в них абзацев с различным предметно-логическим содержанием, можно отметить следующее.

При анализе 25 начальных абзацев выявлено 11 различных по характеру типа для данной позиции абзацев, которые нами условно были обозначены кодовыми значениями от М301 по М311 (см. *табл. 17*). Из них наиболее употребительными оказались:

– "высказывание автора о сути прочего объекта и констатация его состояния в связи с главным субъектом" (М310, f = 0,16);

– "характеристика прочего объекта и высказывание предположниями о будущих действиях прочих субъектов" (М309, f=0,16);

– "описание природной или внешней обстановки, связанной с прочими субъектами и объектами" (М307, f = 0,12) и т.д.

В качестве примера рассмотрим начальный абзац газетной статьи ***"Мүмкіндік кеңінен пайдаланылуда"***: *"Партия программасының жаңа редакциясы жобасында белгіленген халық ағарту саласындағы негізгі міндеттер әрбір педогогтен идеялық-саяси, кәсіптік даярлығын үздіксіз жетілдіріп отыруды талап етеді. Бүгінгі таңда мұғалімдерге педогогикамен методика төңірегіндегі жалпылама шолу сипаттамалар емес, ғылыммен, практикамен ұштастырылған нақты тұжырымдар қажет".*

Таблица 17. **Типы начальных абзацев публицистического текста по их предметно-логическому содержанию**

№ п/п	Типы начальных абзацев	Код типов	Частота	
			F	f
1	2	3	4	5
1	Описание природы и внешней обстановки, связанное с главным объектом	М301	1	0,0400
2	Констатация внешней обстановки, связанной с главными субъектами	М302	1	0,0400
3	Констатация прошлого действия автора и краткая характеристика прочих объектов	М303	3	0,1200
4	Характеристика главного объекта и описание действий автора	М304	1	0,0400
5	Констатация внешней обстановки и действий главного субъекта	М305	3	0,1200
6	Констатация действий автора, связанных с главным субъектом	М306	1	0,0400
7	Описание природной или внешней обстановки, связанной с прочими субъектами и объектами	М307	3	0,1200
8	Констатация некоторого призыва к прочим субъектами	М308	1	0,0400
9	Характеристика прочего объекта и высказывание предположениями о будущих действиях прочих субъектов	М309	4	0,1600
10	Высказывание автора о сути прочего объекта и констатация его состояния в связи с главным субъектом	М310	4	0,1600
11	Констатация некоторой внешней обстановки, связанная с главным объектом	М311	3	0,1200
		Всего	25	1,0000

Рассказанный сюжет в этом абзаце, как нам кажется, больше всего можно отнести к одному из употребительных типов предметно-логического содержания, а именно "описание природной или внешней обстановки, связанной с прочими субъектами и объектами". Здесь к внешней обстановке можно было бы отнести "основные требования проекта новой редакции партийной программы о народном просвещении и т.д., в связи с необходимостью (для учителей) научно обоснованных и проверенных на практике конкретных положений по педогогике и методике преподования".

Можно было бы привести для примера и другие выдержки абзацев по исследуемым газетным текстам, которые свидетельствуют об истинности всех выявленных нами 11 типов предметно-логических содержаний начальных абзацев. Ограничиваясь рассмотренным примером абзаца, перейдем далее к медиальным абзацам публицистического текста, т.е. к абзацам, где изложено основное содержание газетных статей. При анализе медиальных абзацев публицистического текста, согласно данным *табл. 18,* можно утверждать, что выявленные типы в отношении предметно-логического содержания намного превосходят численность начальных и конечных абзацев.

Среди 175 медиальных абзацев публицистического текста нами зафиксирован 41 типовой вариант, относящийся к различным прдметно-логическим содержаниям. Им присвоены кодовые значения в интервале от М401 по М441. Самым частым из всех возможных случаев является тип М413 с относительной частотой 0,0914 (более чем 9% всех типов).

Этот тип имеет формулировку "констатация автором некоторого факта, связанного с прочими субъектами и объектами". Далее следуют типы по мере убывания их частот:

– "констатация некоторых действий, производимых над главным субъектом" (М411, f =0,0743);

– "описание природы или внешней обстановки, связанные с главным объектом" (М403, f =0,0629);

– "констатация действий главного субъекта и характеристика главного объекта" (М422, f =0,0571).

Таблица 18. **Типы медиальных абзацев публицистического текста по их предметно-логическому содержанию**

№ п/п	Типы медиальных абзацев	Код типов	Частота	
			F	f
1	2	3	4	5
1	Констатация прошлого действия автора и краткая характеристика прочих объектов	М401	5	0,0286
2	Краткая характеристика прочего объекта, связанного с второстепенным или главным объектом	М402	7	0,0400
3	Описание природы или внешней обстановки, связанное с главным объектом	М403	11	0,0629
4	Констатация некоторых последовательных действий автора	М404	6	0,0343
5	Описание главного объекта и прочих субъектов	М405	4	0,0229
6	Констатация некоторой обстановки, связанной с главными субъектами	М406	7	0,0400
7	Краткая характеристика главного субъекта	М407	1	0,0057
8	Констатация последовательных действий двух главных субъектов	М408	1	0,0057
9	Констатация некоторых действий главного субъекта	М409	6	0,0343
10	Констатация действий прочего субъекта и состояния главного субъекта	М410	5	0,0286
11	Констатация некоторых действий, производимых над главным субектом	М411	13	0,0743
12	Описание состояния главного субъекта и его действий	М412	6	0,0344
13	Констатация автором некоторого факта, связанного с прочими субъектами и объектами	М413	16	0,0914
14	Описание внешней обстановки и действий прочих субъектов	М414	8	0,0457
15	Констатация прошлых действий автора и характеристика главного субъекта	М415	3	0,0171
16	Описание внутренней обстановки помещения	М416	2	0,0114
17	Характеристика главного субъекта и констатация его последующих действий	М417	4	0,0229
18	Описание внешней обстановки и констатация действий главного субъекта	М418	10	0,0571
19	Констатация действий прочих субъектов	М419	2	0,0114
20	Констатация действий автора и некоторой внешней обстановки	М420	1	0,0057
21	Описание действий второстепенного субъекта	М421	1	0,0057

Окончание табл. 18

1	2	3	4	5
22	Констатация действий главного субъекта и характеристика главного объекта	М422	10	0,0571
23	Констатация некоторого предположения автора, связанного с действиями или состояниями главных субъектов	М423	6	0,0344
24	Констатация действий главного объекта	М424	2	0,0114
25	Развернутое рассуждение автора, связанное с главными и второстепенными объектами и субъектами	М425	5	0,0286
26	Констатация состояния главного субъекта в связи с действиями прочих субъектов	М426	2	0,0114
27	Описание последовательных действий главного субъекта с характеристикой прочих объектов	М427	6	0,0344
28	Описание последовательных действий главного субъекта и описание состояний и действий второстепенного субъекта	М428	4	0,0229
29	Констатация действий и состояния прочих субъектов	М429	2	0,0114
30	Констатация действий второстепенного субъекта с постановкой вопроса	М430	2	0,0114
31	Констатация ответа главного субъекта на предварительные вопросы и постановка им нового вопроса	М431	1	0,0057
32	Констатация ответа главного субъекта на предыдущий вопрос	М432	3	0,0171
33	Описание состояний главного субъекта, действий главного объекта и другого главного субъекта	М433	2	0,0114
34	Описание (характеристика) главного объекта и констатация действий главных субъектов	М434	3	0,0171
35	Констатация действия главного объекта, состояния и будущего действия главного субъекта	М435	1	0,0057
36	Констатация действий главного субъекта и описание внешней обстановки и состояния другого главного субъекта	М436	2	0,0114
37	Констатация действий второстепенных субъектов и характеристика главного объекта	М437	1	0,0057
38	Констатация действий второстепенного субъекта и характеристика и прочего объекта	М438	1	0,0057
39	Констатация действий прочих субъектов и объектов	М439	1	0,0057
40	Констатация последовательных действий прочих и предстоящих действий главного субъекта	М440	1	0,0057
41	Констатация действия автора и описание главного субъекта	М441	1	0,0057
		Всего	175	1,0000

Ниже приводим пример абзаца по тексту (П11-12) из статьи ***"Білімді іске асыру"*** *("Қазақстан мұғалімі".* 1988, N5), который нами отнесен к типу с кодом М422: *"1983 жылдан биофизика бірлестігінің жаңа нұсқаулары бойынша оқушылар тұқымды қоймадағы жерінде сәулелендіру жұмыстарын жүргізе бастады. Тұқымды лазер қондырғысымен өңдеудің нәтижесінде 1984 жылы дәнді дақылдардың әр гектарынан 12,3 центнерден өнім жиналды".*

Согласно содержанию, главным субъектом микроситуации является *"оқушылар" (ученики),* а объектами - *"тұқым"* (т.е. зерно, предназначенное для посева) и *"өнім" (урожай).*

По содержанию абзаца идет констатация действий главного субъекта, т.е. действий учеников, которые облучали лазерным лучом зерно в амбаре. Кроме

того, дается и характеристика главного объекта (зерна), урожай которого с каждого гектара составил 12,3 ц/га.

В качестве конечных абзацев в публицистических текстах нами рассмотрено 45 абзацев, которые, на наш взгляд, являются заключительными или конечными абзацами в 35 газетных статьях. Анализируя ранее известным методом, использованным в классификации научно-популярного текста, мы обнаружили 11 типовых абзацев, относящихся к предметно-логическому содержанию (*табл. 19*).

Более употребительными из 11 типов можно отметить следующие:

– "высказывание автором предположения, связанного с главным субъектом" (М510, f=0,2889);

–"развернутое рассуждение автора, связанное с главными и второстепенными объектами и субъектами" (М508, f=0,2002);

– "констатация заключительного утверждение автора о рассказанном" (М504, f=0,1778).

Например, текст статьи *"Балабақшада оқиды"* в нашем эксперименте имеет обозначение (П20-33) и заканчивается следующим абзацем: *"Алтыжастағыларды оқытуда тек мектепке дейінгі мекемелерге арқа сүйеп жайбарақаттық танытпай мектептердің де мүмкүндіктерін молайту қажет. Бұл мәселе облыстық және қалалық халық ағарту бөлімдері басшыларынан барынша жауапкершілікті талап етеді".*

По всей видимости, содержание этого абзаца можно отнести к предметно-логическому содержанию, относящиемуся к типу "развернутое рассуждение автора, связанное с главными и второстепенными объектами и субъектами", так как, кроме главного объекта (дошкольные учреждение), затрагиваются и возможности школ, и возможности областных и городских отделов народного просвещения, и их руководителей (объекты и субъекты).

Таблица 19. **Типы конечных абзацев публицистического текста по их предметно-логическому содержанию**

№ п/п	Типы конечных абзацев	Код типов	Частота	
			F	f
1	2	3	4	5
1	Констатация утверждения о сути некоторого прочего объекта	М501	4	0,0889
2	Констатация утверждения о сути некоторого главного объекта	М502	2	0,0444
3	Констатация некоторого предположения автора, связанного с действиями или состояниями главных субъектов	М503	1	0,0222
4	Констатация заключительного утверждения автора о рассказанном	М504	8	0,1778
5	Констатация будущих действий автора	М505	1	0,0222
6	Констатация действий автора и характеристика прочего объекта, связанных с главным субъектом	М506	2	0,0444
7	Констатация прошлого действия автора и характеристика прочего объекта	М507	2	0,0444
8	Развернутое рассуждение автора, связанное с главными и второстепенными объектами и субъектами	М508	9	0,2002
9	Констатация природной обстановки и некоторого факта	М509	1	0,0222
10	Высказывание автором предположения, связанного с главным субъектом	М510	13	0,2889
11	Описание состояния автора, связанного с характеристикой главного субъекта	М511	2	0,0444
		Всего	45	1,0000

"Возмем в качестве другого примера заключительной абзац по тексту *(П14-10): Қорыта келгенде, педогогикалық коллектив күнделікті жұмысында мектеп реформасының талаптарын басшылыққа ала отырып, творчестволықпен жұмыс істей беретініне сеніміміз мол".*

Этот абзац, бесспорно, относится к типу "высказывание автором предположения, связанного с главным субъектом", так как автор статьи надеется, что педколлектив будет относиться к своим обязанностям творчески. При сравнении между собой всех абзацев, использованных при создании публицистических (газетных) текстов можно отметить, что, в отличие от научно-популярных текстов, здесь существует определенная специфика в построении начальной, центральной и заключительной части таких текстов. Наряду с этим замечается и общность как в структурном, так и в логическом планах, которая составляет незначительный процент от числа всех абзацев, использованных при создании начальной, центральной и конечной части текстов.

С учетом того, что существует определенная специфика в организации различных частей газетного текста, можно полагать, что научно-популярный текст в этом отношении является более однородным.

4.7. Особенности абзацев художественного текста по их предметно-логическому содержанию

Принципы, использованные в пунктах 4.5 и 4.6 для выделения предметно-логического содержания абзацев, применялись нами и для характеристики абзацев 20 художественных текстов. Предметно-логическое содержание абзацев художественных текстов проанализировано нами без учета позиций в тексте, т.е. мы не разделяли их на начальные, медиальные (центральные) и заключительные. Следует напомнить, что такое разделение художественных текстов было осуществлено при статистическом исследовании их на предмет функционально-смысловой классификации (см. пункт 4.4).

В целях выяснения особенностей абзацев художественного текста по их предметно-логическому содержанию мы рассмотрели 487 абзацев. Особое внимание было обращено на содержание каждого абзаца, при этом учитывалась роль главных и второстепенных объектов и субъектов микроситуации. В результате такого трудоемкого исследования нам удалось выделить 32 типовых по содержанию абзацев, формулировка и статистика которых представлены в *табл. 20.*

Таблица 20. **Типы абзацев художественного текста по их предметно-логическому содержанию**

№ п/п	Типы абзацев	Код типов	Частота	
			F	f
1	2	3	4	5
1	Описание природы или внешней обстановки и констатация некоторого действия главного субъекта	М601	54	0,1109
2	Констатация состояния или действия автора, связанных с главным субъектом	М602	21	0,0431
3	Описание состояния автора	М603	25	0,0513
4	Описание природы и констатация действия или состояния автора	М604	31	0,0636
5	Описание действий главного субъекта, связанных с главным объектом и другим главным субъектом	М605	40	0,0821

Окончание табл. 20

1	2	3	4	5
6	Характеристика главного субъекта, констатация отношения автора к главному субъекту	M606	3	0,0062
7	Описание природы	M607	10	0,0205
8	Описание состояния и действий автора	M608	3	0,0062
9	Констатация будущего действия автора	M609	40	0,0821
10	Характеристика главного объекта и констатация его действий	M610	13	0,0267
11	Констатация будущих действий или состояний автора и прочих субъектов	M611	17	0,0349
12	Описание внешнести, состояния и действий автора, связанных с прочими объектами	M612	13	0,0267
13	Описание внешней обстановки	M613	10	0,0205
14	Описание главного объекта и состояния автора	M614	27	0,0554
15	Описание действий автора и главного объекта	M615	10	0,0205
16	Описание природы и констатация обращения автора к второстепен-ному объекту	M616	4	0,0082
17	Описание состояния автора, природы и будущего действия главного субъекта	M617	3	0,0062
18	Побуждение автора к выполнению действий прочими субъектами или объектами	M618	10	0,0205
19	Характеристика главного субъекта и констатация действий второстепенного объекта	M619	2	0,0041
20	Рассуждение автора, связанное с прочими объектами, и характеристика автора	M620	17	0,0349
21	Описание второстепенного объекта и характеристика автора	M621	10	0,0205
22	Характеристика второстепенного объекта и описание состояний или действий автора	M622	10	0,0205
23	Побуждение автора к выполнению действия прочим субъектом и описание состояния автора	M623	17	0,0349
24	Характеристика автора и констатация его действий	M624	23	0,0472
25	Констатация действия второстепенного субъекта и характеристика автора	M625	20	0,0411
26	Побуждения автора к выполнению действий главным субъектом и констатация некоторых действий и состояния автора	M626	31	0,0636
27	Побуждение автора к выполнению действий главным субъектом и констатация действий другого главного субъекта	M627	20	0,0411
28	Констатация действия или состояния автора и второстепенного субъекта	M628	10	0,0205
29	Характеристика автора и описание действий, выполняемых над автором	M629	13	0,0267
30	Описание природы и констатация некоторого действия второстепенного субъекта, связанного с главным объектом	M630	3	0,0062
31	Констатация действий и состояний второстепенных объектов, а также некоторых действий прочих субъектов, связанных с автором	M631	7	0,0144
32	Констатация состояния и действий автора и побуждение автора к выполнению действий второстепенного объекта	M632	7	0,0144
		Всего	487	1,0000

Как видно из *табл. 20* самыми употребительные среди 32 типовых абзацев оказались следующие типы:

– "описание природы или внешней обстановки и констатация некоторого действия главного субъекта" (М601, f=0,1109);

– "описание действий главного субъекта, связанных с главным объектом и другим главным субъектом" (М605, f=0,0821);

– "описание природы и констатация действия или состояния автора (М604, f=0,0636) и др.

В качестве примера рассмотрим начальный абзац (Х4-01) из рассказа С.Сейфуллина ***"Бандыны қуған Хамит"***: *"Күн шайдай ашық, жылы еді. Аспанда бұлт жоқ, жел де жоқ. Айнала қалың иірім-иірім орман. Бойлап, таласа, жарыса шыққан қайың мен қарағай. Және әредік қалың қайың, қарағаймен жарыса шыққан теректер де көрінеді. Қайың мен терек ағаштардың жапырақтары салғырт тартып, сарғая бастаған. Осы орман ағаштардың ішімен, ескілеу жолмен Хамит жүріп келеді".*

Итак, этот начальный абзац художественного текста по функционально-смысловому характеру относится к описательному абзацу, а по предметно-логическому содержанию к описанию природы (живописного леса) и действий главного субъекта (продвижение Хамита по лесной тропинке), т.е. этот абзац, относится по содержанию к типу, который имеет код М601.

Проанализировав данные *Табл. 20* отметим, что для художественного текста свойственны типы, относящиеся к описанию внешней среды, состояния или действия главных персонажей, а также характеристики главного субъекта, объекта и констатация будущих действий автора или прочих субъектов и т.д. Эти тексты значительно отличаются по своему содержанию от публицистических текстов, и, тем более, от научно-популярного текста.

В заключении этой главы отметим, что названные типы классов абзацев по функционально-смысловому содержанию и по предметно-логическому содержанию могут изменяться, как в количественном, так и в качественном (содержательном) аспектах в связи с характером текстов. Разработанный нами принцип выделения типовых абзацев должен служить основой при формализации содержания казахского текста.

Глава пятая
ОСНОВНОЕ СТАТИЧЕСКОЕ СОДЕРЖАНИЕ ТЕКСТА И ЕГО ФОРМАЛЬНОЕ ОПИСАНИЕ

5.1. Статика и динамика содержания текста

Исследование текстов, проведенных в различных аспектах, характеризует текст как многоплановое образование. Такое утверждение обосновано на том факте, что в любом тексте можно обнаружить социальное и индивидуальное, детерминированное и случайное, глубинное и поверхностное, а также языковое и речевое [237, с.43]. К таким же противоречивым явлениям относится наличие в текстах статических и динамическких аспектов, которые следует понимать в смысле "покой" и "движение" [70, с.19]. Текст в процессе его порождения, восприятия и понимания считается находящимся в движении или динамическом состоянии. В остальных случаях текст, как последовательность лексических единиц, как результат, продукт речемыслительной деятельности, нужно полагать, находится в статическом состоянии или в состоянии покоя [70, с.19; 209, с.31].

Известно, что в процессе воспроизведении предложений, прежде всего, запоминаются такие сведения о предложении, как информация об его семантическом содержании и синтаксической структуре. Причем, семантический аспект в процессе воспроизведения всегда предшествует синтаксическому [175, с.89].

В последнее время лингвисты стали больше применять термины "словарь" и "синтаксис" по отношению к тексту, констатируя о том, что каждый текст должен иметь свой словарь и свой синтаксис [153, с.123; 152, с.141; 237, с.56]. С учетом того, что терминам "словарь" и "синтаксис" исследователи придают разный смысл, мы будем придерживаться такой точки зрения, когда под словарем понимается слова "содержание" или "несомые слова" вместе с "несущими" или служебными словами [237, с.44].

Таким образом, с нашей точки зрения, статику текста следует понимать как слова "содержания", которые служат в тексте отражением некоторого множества предметов, явлений, фактов реальной действительности, а динамику – как синтаксис, отражающий отношения между элементами указанного множества. Эти отношения устанавливаются автором текста в зависимости от целевой установки, типа текста, речевого опыта и других факторов. О том, что в тексте наличествуют статический и динамический аспекты, в указанном нами смысле, подтверждают результаты экспериментов по анализу процесса понимания текста [182, с.120].

Рассмотрим эти две составляющие текста подробнее с точки зрения его организации как единого целого.

Как мы уже упоминали, в тексте отражаются различные ситуации окружающей действительности. Элементы этой ситуации воспринимаются человеком в условиях некоторой вероятности. Такой процесс зависит от многих факторов, связанных с жизненным и языковым опытом воспринимающего индивида [54, с.81; 70, с.40]. В результате происходит отбор узловых элементов системы референтов при развертывании содержания, который является очень существенным в процессе построения текста. Эти узловые элементы дают возможность устанавливать порядок дальнейшего отбора из подмножества связанных с ним референтов последующих

элементов, необходимых для построения текста [209, с.70-71]. Если учесть при этом, что референтами в тексте служат различные имена, то построение формальной структуры содержания текста это и есть формальная структура линейной последовательности материальных знаков, т.е. слов [209, с.28].

Пронизывающими весь текст на всей его протяженности и отражающими его смысловое единство являются так называемые "опорные слова". Эти слова в лингвистических источниках встречаются в различной терминологии: "ключевые слова", "смысловые вехи", "топики", "дескрипторы", "ядра", "фокальные точки" и т.д. Роль "опорных слов" в тексте, характеризующих смысловое единство, показана во многих исследовательских работах по смысловой организации текста [4, с.15; 189, с.28; 248, с.35; 280, с.102; 315, с.1-8; 329]. В этой литературе имеются довольно интересные попытки формального представления смысла текста через такие опорные слова и словосочетания, связанные с построением различного рода графов. Однако, как утверждают авторы подобных построений, такая формальная связь не позволяет полностью решить задачу построения графа денотативной структуры, соответствующего всему тексту. Это объясняется тем, что окончательная связь всех его вершин, представляющих текст, осуществляется на содержательном уровне с привлечением информации, не содержащейся в тексте [209, с. 149-150].

Дифференцированный анализ "опорных слов" показывает, что в их структуре существует определенная иерархия [160, с.1-5; 182, с.123], а именно часть таких "опорных слов" является "главными", "важными словами", "номенклатурными дескрипторами", "ключевыми словами", "выступающими точками" всего содержания текста [57, с.201; 160; 182, с.123; 189, с.28; 209, с.147; 248, с.35]. Можно полагать, что эти слова, определяя главный предмет сообщения, представляют собой как-бы свернутый замысел текста. Следует отметить и наличие в тексте контекстуальных синонимов, т.е. таких слов, когда один и тот же референт может выражаться синонимическими словами [182, с.138]. Из-за линейности означающего текста контекстуальные слова распределяются в тексте в разных абзацах.

Проблема смысловой организации абзаца является в целом семантико-синтаксической [152, с.137], в тоже время синтаксические особенности этой единицы плана содержания не могут быть поняты без детального проникновения в ее лексическую структуру. В абзацах отражаются те связи и отношения, которые зафиксированы у главных и второстепенных референтов текста с другими, менее значимыми элементами ситуации: различные типы действий и состояний ГОС и ВОС, различные места действий, время действий, цели, причины действий и состояний и т.п. [208, с.53-54; 307, с.108].

Второй тип составляющих текста – динамика текста, которая характеризует отношения между объектами реальной действительности, зафиксированные в конкретной ситуации, описанной в тексте. Здесь речь идет о фиксации тех событий (процессов, действий, состояний), которые происходят с референтами [161, с.324], имена которых представлены в тексте опорными словами.

Известно, что членение действительности на фрагменты проводится не произвольно, а в соответствии с социально отработанными моделями, сформировавшимися в ходе длительного исторического развития человечества [305, с.42]. Это может быть объяснено ограниченнстью оперативной памяти человека и невозможностью полного охвата событий. И поэтому человеческое сознание в процессе познания вычленяет некоторый фрагмент событий, что позволит последовательно рассмотреть различные признаки одного предмета при подведении каждого из них под общую категорию, известную ранее и зафиксированную в сознании [305, с.41].

Ученые самых различных направлений отмечают, что в процессе человеческого познания наиболее важные и часто повторяющиеся отношения реальной действительности закрепляются в форме синтаксических структур. Такие синтаксические структуры есть не что иное, как некоторые устоявшиеся, отложившиеся в памяти логические образования, отразившиеся в человеческом мышлении типические связи и отношения объективной действительности [5, с. 103-104]. Другими словами, это есть речевые шаблоны-высказывания или "логические слепки" с действительности. Причем набор таких шаблонов относительно невелик и специфичен для текстов одного автора или одной достаточно узкой предметной области [261, с.297]. Исследования последних лет позволяют допустить, что и абзац письменного текста также является "логическим слепком" с действительности [225]. Это подтверждают не только лингвистические исследования, но и эксперименты психологов и психолингвистов [269, с.156].

5.2. Способы выделения опорных и ключевых слов текста

Проблема выделения основного и факультативного содержания текста является одной из важных в инженерной лингвистике. Из множества существующих способов выделения главных опорных слов (ГОС) наиболее перспективным нам представляется метод, использованный в работах [187, с.82-89; 125, с.75-94]. В указанных работах предлагаются два формальных критерия отделения ГОС, т.е. в его основе лежит идея о том, что такая наиболее значимая единица текста должна иметь, во-первых, наибольшую частоту встречаемости в тексте, во-вторых, встречаться в наибольшем количестве абзацев такого текста. Как показывают результаты подобных исследований, такими опорными словами, как правило, являются имена, отражающие главные действующие лица или главные предметы таких типов текста, как описание, повествование или рассуждение.

Выделенные из анализируемого текста главные опорные слова в процессе развертывания содержания становятся своеобразным центром, вокруг которого формируются другие элементы как составляющие микроситуации. Такими составляющими обычно служат следующие классы слов: глаголы, прилагательные, наречия и т.п. [208, с.53].

В зависимости от ситуации, главные референты (им соответствуют ГОС) могут быть взаимосвязаны с другими референтами, относительно важными для рассматриваемой ситуации. Как правило, последние являются центром некоторых микроситуаций по отношению к общей ситуации [209]. В исследовательской практике слова, соответствующие относительно важным референтам, называют второстепенными опорными словами (ВОС) [126, с.177-178]. Естественно, что ВОС по частотности стоят ниже, чем ГОС, т.е. всречаются с меньшей частотой в тексте, чем ГОС и охватывают меньшее число абзацев. Все остальные слова текста, не относящиеся к ГОС и ВОС, мы назвали прочими словами, т.е. сокращенно ПС.

Выделенные экспериментальным путем ГОС и ВОС текста оформляются в специальную таблицу основного статического содержания (ТОСС).

Известно, что, формируя содержание текста, автор проводит отбор конкретных слов и их распределение по тексту. Причем, это осуществляется таким образом, чтобы потенциальный получатель текста мог сформировать в своем сознании то же представление об описываемом фрагменте ситуации, которое вложил в текст автор. Отбор конкретных слов и включение их в различные типы сочетаний проводятся автором в зависимости от темы текста, т.е. "о ком (чем) будет текст". Под темой мы будем понимать понятийное ядро текста, передающее содержание

текста в сжатом и абстрактном виде, в форме смыслового комплекса, выраженного словесными средствами и состоящего из имен главных действующих лиц и предикатов. Иными словами, тема – это совокупность определенных слов (и словосочетаний), отражающая наиболее существенные составляющие описываемой в тексте ситуации и наиболее важные отношения между этими составляющими [281, с.336]. Это как раз то инвариантное, что содержится в тексте и одинаково выделяется и автором текста, и основной массой воспринимающих текст. Наличие таких инвариантных смысловых единиц подтверждается и психолингвистическими экспериментами по смысловому анализу и синтезу текстов различными испытуемыми [51; 126, с.174].

Интуитивно ясно, что каждый текст представляет собой единое смысловое целое потому, что в нем, во всех его частях говорится "об одном и том же", т.е. на всем протяжении текста действуют одни и те же субъекты и объекты, что события в относительно небольшом законченном тексте происходят в какое-то ограниченное время и в фиксированном числе мест. Но для передачи описываемой ситуации с индивидуальной точки зрения автора, с индивидуальными целями и намерениями такие составляющие могут быть неравноценны для автора. Эта неравноценность приводит к тому, что часть из них действует в большинстве микроситуаций описываемой единой ситуации, другие же – лишь в некоторых из них. Это приводит к тому, что среди имен текстов, представляющих упомянутые составляющие, имеется определенная иерархия, соотносимая с иерарихой соответствующих денотатов [182, с.123; 208, с.55].

Прежде всего, каждая ситуация передается в тексте конкретным автором. Именно его жизненный опыт, его восприятие мира отражены в тексте. Именно автор определяет динамику текста, являясь "эго" всего текста [266, с.332]. Поэтому имена текста, отражающие личность самого автора, будут являться для нас главными опорными словами текста (независимо от частоты употребления). К этой группе слов мы отнесем и те слова текста, которые передают: 1) главные действующие лица; 2) главные объекты действия; 3) основные места действия; 4) времена действия.

Как правило, эти слова имеют в тексте наибольшую частоту употребления и используются в большем числе фрагментов (абзацев) текста, образуя соответствующие "изотопические цепочки", "тематические прогрессии", "номинативные цепи", "рекуррентные цепи", "семантические изотопии", "имплицитные референции", "лексико-тематические линии", "нанизывания" и т.п. [4, с.14; 144, с.121; 199, с.19; 237, с.49]. Основная особенность таких последовательностей лексических единиц текста заключается в том, что они имеют в тексте одну и ту же предметную соотнесенность. В таком ряду могут присутствовать и полные словарные синонимы, и контекстуальные синонимы, и ассоциативные замены, и, наконец, замены местоименные [4, с.15; 173, с.23-25]. В нашу задачу не входит детальный анализ способов всех возможных замен в указанных семантических цепочках. Будем считать, что представленные четыре способа являются способами денотативного отождествления имен.

Существует достаточно большое число способов определения "степени важности" для содержания текста того или иного имени или именного словосочетания [41, с.235-236; 126, с.176]. Их можно разделить на: 1) анкетные; 2) структурные; 3) словарные; 4) частотные; 5) синтаксические; 6) диалоговые.

Все эти методы имеют один существенный недостаток: они не исходят из текста как единой связной единицы. Например, с помощью структурных методов выделяются опорные слова на основе их вхождения в заголовок, первое предложение или нескольких отдельных предложений. Частотные методы исходят из частоты встречаемости слов в тексте. При таком подходе по существу учитывается не

смысл текста, а свойства плана его выражения [126, с.176]. Определенные недостатки существенны и для других способов поиска опорных слов текста.

5.3. Способ выделения опорных слов казахского текста

Как мы отмечали, основными критериями при выделении опорных слов текста являются абсолютная частота употребления слова (с учетом всех возможных замен) и количество абзацев, в которых встретилось слово, кроме этого, такой критерий не должен зависеть от общего числа слов в тексте.

Наиболее удобной для нас представляется несколько измененная формула коэффициента статистической устойчивости термина [187, с.87]:

$$Кваж = \frac{F \bullet m}{N \bullet n} \qquad (1)$$

где F – абсолютная частота слова в тексте (в нее входит суммарная частота всех типов синонимов этого слова и местоименных замен); m – число абзацев, в которых встретилось слово; N – общие число слов в тексте; n – общее число абзацев в тексте.

Назовем этот критерий *Кваж* **коэффициентом важности слова** и определим для слов всех анализируемых текстов его критические значения $K^1важ$ и $K^2важ$, позволяющие формальным способом отделить в массе слов конкретного текста соответственно главные и второстепенные опорные слова.

Эксперименты по выделению опорных и неопорных слов наших трех видов текстов, проведенные на текстах самой различной длины (от одного абзаца до 100), дают нам возможность предложить на основе формулы (1) экспериментальным путем составленные нами две формулы-неравенства (2) и (3):

$$\frac{\left(\frac{n}{m_1}+1\right)^2}{N \cdot n} \leq K^1 \text{важ} \leq 1 \qquad (2)$$

$$\frac{\left(\frac{n}{m_2}+1\right)^2}{N \cdot n} \leq K^2 \text{важ} < \frac{\left(\frac{n}{m_1}+1\right)^2}{N \cdot n} \qquad (3)$$

С помощью этих неравенств можно определить левые и правые границы отнесения потенциальных слов к ГОС и ВОС для текстов на казахском языке.

Здесь неравенство (2) служит для определения границ ГОС, а (3) – для ВОС. Значения коэффициентов m_1 и m_2, участвующих в формулах, зависят от общего числа абзацев в тексте и поэтому они были нами подобраны экспериментальным путем. Значения m_1 и m_2 будут зависеть от интервалов, в которые попадет значение n – общее число абзацев в тексте. Например, для K^1важ (т.е. для ГОС), если общее число абзацев n более 100 или равно 100, то $m_1 = 42$, а если n более 60, но менее 100, то $m_1 = 30$ и т.д. Полный перечень значений для m_1 и m_2 (для ГОС и ВОС) и соответствующие им формулы неравенств для K^1важ и K^2важ приведены в *табл. 21* и в *табл. 22.*

Таким образом, можно предложить следующие критерии:

1) к **главным опорным словам** (ГОС) текста будем относить те слова, которые удовлетворяют требованиям, определенным в *табл. 21;*

2) к **второстепеным опорным словам** (ВОС) того же текста относятся те его слова, которые удовлетворяют требованиям, определенным в *табл. 22;*

3) все оставшиеся слова текста считаем неопорными словами и будем называть их **прочими словами** (ПС).

В то же время слова, входящие в каждую из первых двух полученных групп, неоднородны по содержанию. В соответствии с предметными свойствами референтов они образуют группы опорных слов – субъектов, объектов, слов-мест и слов-времен [73, с.23; 183, с.15-51; 307, с.105]. Именно они совместно с предикатами выражают взаимосвязи основных, объективно существующих, категорий: материи, движения, времени и пространства [183, с.15-51].

Таблица 21. **Критерии отнесения слов казахского текста к главным опорным словам**

№ п/п	Число абзацев в тексте: n	Число абзацев, в которых встретилось слово: m_1	Пределы изменения K^1важ для отнесения потенциальных слов к ГОС
1	2	3	4
1	$n \geq 100$	m_1=42	$\frac{n}{(\frac{42}{}+1)^2}.\frac{1}{N \cdot n} \leq K^1$важ< 1
2	$81 \leq n \leq 99$	m_1=30	$\frac{n}{(\frac{30}{}+1)^2}.\frac{1}{N \cdot n} \leq K^1$важ<1
3	$51 \leq n \leq 80$	m_1=17	$\frac{n}{(\frac{17}{}+1)^2}.\frac{1}{N \cdot n} \leq K^1$важ<1
4	$31 \leq n \leq 50$	m_1=16	$\frac{n}{(\frac{16}{}+1)^2}.\frac{1}{N \cdot n} \leq K^1$важ<1
5	$15 \leq n \leq 30$	m_1=12	$\frac{n}{(\frac{12}{}+1)^2}.\frac{1}{N \cdot n} \leq K^1$важ<1
6	$9 \leq n \leq 14$	m_1=5	$\frac{n}{(\frac{5}{}+1)^2}.\frac{1}{N \cdot n} \leq K^1$важ<1
7	$6 \leq n \leq 8$	m_1=4	$\frac{n}{(\frac{4}{}+1)^2}.\frac{1}{N \cdot n} \leq K^1$важ<1
8	$n=4; n=5$	m_1=3	$\frac{9}{N \cdot n} \leq K^1$важ<1
9	$n=3$	m_1=2	$\frac{4}{N \cdot n} \leq K^1$важ<1
10	$n=1; n=2$	m_1=1	$\frac{3}{N \cdot n} \leq K^1$важ<1

Таблица 22. **Критерии относения слов казахского текста к второстепенным опорным словам**

№ п/п	Число абзацев в тексте: n	Число абзацев, в которых встретилось слово: m_2	Пределы изменения K^2важ для отнесения потенциальных слов к ВОС
1	2	3	4
1	$n \geq 100$	m_2=80	$\frac{n}{(80+1)^2}.\frac{1}{N \cdot n} \leq K^2важ < K^1важ$
2	$81 \leq n \leq 99$	m_2=64	$\frac{n}{(64+1)^2}.\frac{1}{N \cdot n} \leq K^2важ < K^1важ$
3	$51 \leq n \leq 80$	m_2=38	$\frac{n}{(38+1)^2}.\frac{1}{N \cdot n} \leq K^2важ < K^1важ$
4	$31 \leq n \leq 50$	m_2=36	$\frac{n}{(36+1)^2}.\frac{1}{N \cdot n} \leq K^2важ < K^1важ$
5	$15 \leq n \leq 30$	m_2=30	$\frac{n}{(30+1)^2}.\frac{1}{N \cdot n} \leq K^2важ < K^1важ$
6	$9 \leq n \leq 14$	m_2=14	$\frac{n}{(14+1)^2}.\frac{1}{N \cdot n} \leq K^2важ < K^1важ$
7	$6 \leq n \leq 8$	m_2=8	$\frac{n}{(8+1)^2}.\frac{1}{N \cdot n} \leq K^2важ < K^1важ$
8	$n=4; n=5$	m_2=2	$\frac{4}{N \cdot n} \leq K^2важ < K^1важ$
9	$n=3$	m_2=1	$\frac{3}{N \cdot n} \leq K^2важ < K^1важ$
10	$n=1; n=2$	m_2=1	$\frac{2}{N \cdot n} \leq K^2важ < K^1важ$

5.4. Выделение опорных слов в трех анализируемых типах текстов

Как видно из предыдущего пункта, для того чтобы выделить опорные слова конкретного текста, необходимо знать общую частоту употребления слова в тексте и распределение этого слова по абзацам текста. Чтобы получить эти данные для каждого текста рассматриваемых нами трех функциональных типов текстов, необходимо составить так называемый распределительный алфавитно-частотный словарь употреблений слов по абзацам. В этом словаре приводятся алфавитные списки всех слов, встретившихся в каждом абзаце, с указанием частоты употребления каждого слова в этом абзаце. Далее проводится объединение словарей абзацев в единый алфавитный словарь текста. При таком объединении суммируются частоты одинаковых слов и подсчитывается число абзацев, в которых встретилось каждое слово. При объединении слов учитываются отмеченные четыре способа замен в тексте: словарный

синонимический, контекстуальный синонимический, местоименный и ассоциативный. Все это делается с опорой на содержание каждого текста.

На следующем шаге анализа из полученного алфавитно-частотного словаря отдельно выделяются имена существительные (и именные словосочетаниях) и другие слова, относящиеся к знаменательным классам слов казахского языка, имеющие частоту *F,* равную двум и более, с указанием всех возможных замен в тексте и числа абзацев, в которых встретились эти слова. Такие слова располагаются в порядке убывания их абсолютных частот *F* и образуют потенциальный словарь опорных слов. Этот словарь служит основой для разделения опорных слов на главные и второстепенные и выделения внутри них различных подгрупп опорных слов.

В нашем эксперименте были рассмотрены три вида текстов по функционально-смыловому типу: 1) научно-популярный стиль – изучены 47 различных по объему (по количеству слов) минитекстов по одной тематике; 2) публицистический стиль (газетные статьи) – 35 газетных статей по одной рубрике (или 35 различных по объему мини-текстов); 3) художественный стиль – 20 мини-текстов различного объема (рассказы видных казахских писателей).

По изложенной последвательности на основе специально составленных нами компьютерных программ были получены алфавитно-частотный распределительный по абзацам словарь по отдельным текстам каждого типа и объединенный словарь по трем типам казахских текстов (Во второй главе мы уже приводили перечень полученных на компьютере словарей. Фрагменты этих словарей будут показаны в Приложении.)

Далее из этих распределительных словарей были получены так называемые потенциальные словари, которые являются базовыми словарями для определения ГОС, ВОС и прочих слов рассматриваемых текстов. Для отдельного типа текстов и входящих в него мини-текстов по известному общему числу слов в тексте *N,* общему числу абзацев в каждом мини-тексте *n* заранее были вычислены пределы изменения $K^1_{важ}$ и $K^2_{важ}$, которые приведены в *табл. 23* для научно-популярных текстов, в *табл. 24* – публицистического и в *табл. 25* – художественного текста.

Для того чтобы читатель имел некоторое представление о таких словарях, приведем фрагмент (отрывок) потенциального словаря в виде *табл. 26,* рассматривая XUD-03 (художественный) и слова, относящиеся к классу существительных.

В потенциальных распределительных алфавитно-частотных словарях для каждого слова нами вычислены значения $K_{важ}$. Сравнением $K_{важ}$ со значениями $K^1_{важ}$ и $K^2_{важ}$ из *табл. 23, 24, 25* (согласно типу текста), установлена их принадлежность или к ГОС, или к ВОС, или к ПС. Эти сведения записаны в последней графе *табл. 26* (фрагмент потенциального словаря по XUD-03).

Поясним (для филологов), как определяется принадлежность слова к опорным словам, согласно предлагаемой нами методике.

Для примера рассмотрим слово ***"ауыл" (село)***, из потенциального словаря, составленного по тексту XUD-03, *N=361, n=11,* (*табл. 25* и *26).*

Согласно формуле (1), $K_{важ}=Fx\ m/Nx\ n$; $Fx\ m=5x4=20$, $Nxn=361x11=3971$, то $K_{важ}=20/3971=0,00503$. Здесь данные *N=361, n=11* и *Nx n=3971* взяты из *табл. 25,* а значения абсолютной частоты слова *F=5, m=4* заимствованы из потенциального словаря, т.е. из *табл. 26.* Вычисленное нами значение $K_{важ}=0,00503$ также отражено в этой *табл.* Теперь, чтобы слово *"ауыл"* отнести к ГОС, ВОС или ПС, необходимо воспользоваться формулами (2) и (3), с помощью которых можно вычислить левую и правую границы интервалов ГОС и ВОС. Ввиду простоты вычисления значения $K^1_{важ}$ и $K^2_{важ}$ по формулам (2) и (3), мы решили воспользоваться их готовыми значениями из *табл. 25,* которые для текста XUD-03 соответственно равны 0,0026 и 0,0008. Как мы убедились, значение $K_{важ}=0,00503$ и оно больше, чем $K^1_{важ}=0,0026$, но меньше чем 1, а это удовлетворяет условию, когда слово из текста XUD-03 *"ауыл"* должно принадлежать к ГОС.

Также можно убедиться, что и слово из этого же текста (XUD-03, N=361, n=11) *"қазақ"* (F=3, m=3) относится к ВОС, а слово *"комитет"* (F=3 и m=1) относится не к опорным, а к прочим словам (ПС).

Таблица 23. **Пределы изменения значений $K^1_{важ}$ и $K^2_{важ}$ для научно-популярного типа казахских текстов**

№ п/п	Имя текста (NAU-ij)	Число слов N	Число абзацев n	N×n	$K^1_{важ}$: от K^1 до 1 K^1	$K^2_{важ}$: от K^1 до K^2 K^2
1	2	3	4	5	6	7
1	NAU-01	188	5	940	0,00957	0,00426
2	NAU-02	750	16	12000	0,00045	0,00019
3	NAU-03	49	2	98	0,03061	0,02041
4	NAU-04	435	7	3045	0,00248	0,00115
5	NAU-05	523	12	6276	0,00184	0,00055
6	NAU-06	329	8	2632	0,00342	0,00152
7	NAU-07	199	3	597	0,00670	0,00503
8	NAU-08	484	8	3872	0,00232	0,00103
9	NAU-09	102	3	306	0,01307	0,00980
10	NAU-10	594	14	8316	0,00174	0,00048
11	NAU-11	71	3	213	0,01878	0,01408
12	NAU-12	861	15	12915	0,00039	0,00017
13	NAU-13	253	4	1012	0,00889	0,00395
14	NAU-14	569	11	6259	0,00164	0,00051
15	NAU-15	411	7	2877	0,00263	0,00124
16	NAU-16	292	5	1460	0,00616	0,00274
17	NAU-17	125	3	375	0,01067	0,00800
18	NAU-18	251	5	1255	0,00717	0,00319
19	NAU-19	280	4	1120	0,00804	0,00357
20	NAU-20	322	8	2576	0,00349	0,00155
21	NAU-21	250	4	1000	0,00900	0,00400
22	NAU-22	937	17	15929	0,00036	0,00016
23	NAU-23	279	3	837	0,00478	0,00358
24	NAU-24	336	6	2016	0,00310	0,00152
25	NAU-25	524	9	4716	0,00166	0,00056
26	NAU-26	806	17	13702	0,00042	0,00018
27	NAU-27	221	4	884	0,01018	0,00452
28	NAU-28	242	5	1210	0,00744	0,00331
29	NAU-29	742	14	10388	0,00139	0,00039
30	NAU-30	343	10	3430	0,00262	0,00085
31	NAU-31	1382	25	34550	0,00028	0,00010
32	NAU-32	1112	18	20016	0,00031	0,00013
33	NAU-33	1800	46	82800	0,00018	0,00006
34	NAU-34	1086	25	27150	0,00035	0,00012
35	NAU-35	877	20	17540	0,00041	0,00016
36	NAU-36	967	22	21274	0,00037	0,00014
37	NAU-37	52	2	104	0,02885	0,01923
38	NAU-38	135	4	540	0,01667	0,00741
39	NAU-39	150	3	450	0,00889	0,00667
40	NAU-40	117	4	468	0,01923	0,00855
41	NAU-41	738	24	17712	0,00051	0,00018
42	NAU-42	720	19	13680	0,00049	0,00019
43	NAU-43	975	21	20475	0,00037	0,00014
44	NAU-44	918	21	19278	0,00039	0,00015
45	NAU-45	524	12	6288	0,00184	0,00053
46	NAU-46	1339	37	49543	0,00022	0,00008
47	NAU-47	589	19	11191	0,00060	0,00023

Таблица 24. **Пределы изменения значений $K^1_{важ.}$ и $K^2_{важ.}$ для публицистического типа казахских текстов**

№ п/п	Имя текста (PUB-ij)	Число слов N	Число абзацев n	N×n	$K^1_{важ}$: от K^1 до 1 K^1	$K^2_{важ}$: от K^1 до K^2 K^2
1	2	3	4	5	6	7
1	PUB-01	358	5	1790	0,00503	0,00223
2	PUB-02	995	20	19900	0,00036	0,00014
3	PUB-03	1088	24	26112	0,00345	0,00124
4	PUB-04	1010	19	19190	0,00035	0,00014
5	PUB-05	902	13	11726	0,00111	0,00031
6	PUB-06	198	6	1188	0,00526	0,00258
7	PUB-07	288	5	1440	0,00625	0,00278
8	PUB-08	485	9	4365	0,00180	0,00060
9	PUB-09	956	26	24856	0,00041	0,00014
10	PUB-10	1141	24	27384	0,00033	0,00012
11	PUB-11	576	14	8064	0,00179	0,00050
12	PUB-12	410	7	2870	0,00264	0,00122
13	PUB-13	410	8	3280	0,00274	0,00122
14	PUB-14	389	11	4279	0,00239	0,00075
15	PUB-15	337	7	2359	0,00321	0,00095
16	PUB-16	374	5	1870	0,00481	0,00214
17	PUB-17	143	3	429	0,00932	0,00699
18	PUB-18	289	8	2312	0,00389	0,00173
19	PUB-19	1061	22	23342	0,00034	0,00013
20	PUB-20	951	30	28530	0,00043	0,00014
21	PUB-21	816	27	22032	0,00048	0,00016
22	PUB-22	822	10	8220	0,00109	0,00035
23	PUB-23	724	14	10136	0,00142	0,00039
24	PUB-24	689	14	9646	0,00150	0,00041
25	PUB-25	979	28	27412	0,00040	0,00013
26	PUB-26	562	13	7306	0,00177	0,00050
27	PUB-27	509	18	9162	0,00068	0,00028
28	PUB-28	454	13	5902	0,00219	0,00062
29	PUB-29	380	9	3420	0,00299	0,00077
30	PUB-30	510	8	4080	0,00221	0,00098
31	PUB-31	754	31	23374	0,00036	0,00015
32	PUB-32	450	12	5400	0,00214	0,00065
33	PUB-33	454	12	5448	0,00212	0,00065
34	PUB-34	481	14	6734	0,00214	0,00059
35	PUB-35	792	16	12672	0,00042	0,00019

Таблица 25. **Пределы изменения значений $K^1_{важ.}$ и $K^2_{важ.}$ для художественного типа казахских текстов**

№ п/п	Имя текста (PUB-ij)	Число слов N	Число абзацев n	N×n	$K^1_{важ}$: от K^1 до 1 K^1	$K^2_{важ}$: от K^1 до K^2 K^2
1	2	3	4	5	6	7
1	XUD-01	274	4	1096	0,0082	0,0036
2	XUD-02	370	13	4810	0,0027	0,0008
3	XUD-03	361	11	3971	0,0026	0,0008
4	XUD-04	2432	74	179968	0,0002	0,00005
5	XUD-05	753	43	32379	0,0004	0,00015
6	XUD-06	1007	54	54378	0,0003	0,00011
7	XUD-07	347	9	3123	0,0025	0,00086
8	XUD-08	150	4	600	0,0150	0,00670
9	XUD-09	473	17	8041	0,00073	0,00031
10	XUD-10	1121	33	36993	0,00025	0,00010
11	XUD-11	988	33	32604	0,00029	0,00011
12	XUD-12	218	15	3270	0,00155	0,00069
13	XUD-13	150	8	1200	0,00750	0,00330
14	XUD-14	577	24	13878	0,00065	0,00023
15	XUD-15	1144	37	42328	0,00026	0,000097
16	XUD-16	1233	37	45621	0,00024	0,00009
17	XUD-17	908	18	16344	0,00038	0,000157
18	XUD-18	115	7	805	0,00940	0,00437
19	XUD-19	173	9	1557	0,00504	0,00173
20	XUD-20	572	37	21164	0,00052	0,00019

Таблица 26. **Потенциальный алфавитно-частотный распределительный словарь художественного текста XUD-03** ***(фрагмент)***

XUD-03: $K^1_{важ.} \geq 0{,}0026$; $K^2_{важ.} \geq 0{,}0008$					
№ п/п	Слово	Абсолютная частота $F \geq 2$	Число абзацев m	$K_{важ.}$	ГОС, ВОС, ПС
1	Ақан/ат	5	3	0,00377	ГОС
2	Ақмола/жа	2	2	0,00100	ВОС
3	ара/зт	2	2	0,00100	ВОС
4	ауыл/зт	5	4	0,00503	ГОС
5	банды/зт	2	2	0,00100	ВОС
6	ел/зт	2	2	0,00100	ВОС
7	ет/зт	2	2	0,00100	ВОС
8	жолдас/зт	3	2	0,00151	ВОС
9	жігіт/зт	2	2	0,00100	ВОС
10	комитет/зт	3	1	0,00075	ПС
11	Кудря/ат	6	3	0,00453	ГОС
12	кісі/зт	2	2	0,00100	ВОС
13	қазақ/зт	3	3	0,00226	ВОС
14	қас/зт	2	2	0,00100	ВОС

Индексы в конце слов (после наклонной черты – /), означают принадлежность слова к классу слов. В *табл. 26:* "ат" сокращение от *адам аты* (имена людей), "жа" – *жер аты* (название местности) и "зт" – *зат есім* (существительное).

5.5. Построение таблицы основного статического содержания текста

Известно, что при задании содержания текста должно быть указано *кто (*или *что)* совершает события, над *чем (кем)* совершаются они, *где и в какое время,* т.е. это есть не что иное, как отражение основных категорий диалектического материализма – материи, движения, времени и пространства [183, с.15-42]. Поэтому среди опорных слов текста (ГОС и ВОС) можно особо выделить и записать в виде "Таблицы основного статического содержания" такие типы, которые относятся к словам-субъектам, словам-объектам, а также словам, означающим место и время происхождения события в зависимости от их десигнативной отнесенности. В *табл. 27* конкретизированы типы опорных слов и каждому типу присвоен буквенно-цифровой код. Цифровые индексы могут изменяться от 01 до 99 в зависимости от количества слов, отнесенных к данному типу.

Таблица 27. **Таблица основного статического содержания (ТОСС)**

Группа опорных слов	Тип опорного слова	Код
Главные опорные слова (ГОС)	Слова-субъекты	Cij
	Слова-объекты	Kij
	Слова, обозначающие место	Mij
	Слова, обозначающие время	Tij
Второстепенные опорные слова (ВОС)	Слова-субъекты	Sij
	Слова-объекты	Gij
	Слова, обозначающие место	Pij
	Слова, обозначающие время	Bij

По потенциальному словарю каждого текста можно выделить главные и второстепенные опорные слова, и согласно содержанию рассматриваемого текста, отнести их к соответствующим типам, указанным в *табл. 27.*

В качестве примера рассмотрим по одному тексту из каждого рассматриваемого нами стилей (XUD-03, PUB-06, NAU-15) и покажем, как конкретизируются опорные слова в виде *табл. 28, 29* и *30* "Основного статического содержания" (сами тексты и другие фрагменты потенциальных словарей на каждый стиль будут приведены в Приложений).

Создавая тот или иной текст, автор представляет перед собой реально (или мысленно) некоторую ситуацию. Он может выделять фрагменты этих ситуаций, а в пределах фрагментов – элементы ситуации со всеми относящимися к этим элементам признаками в их связи с другими элементами той же ситуации. Это и есть энциклопедические знания, и потому проблема представления энциклопедических знаний и проблема организации словаря текста должны рассматриваться в неразрывном единстве [180, с.99].

Таблица 28. **Таблица основного статического содержания (ТОСС по XUD-03)**

Конкретное значение опорных слов	Группа опорных слов			
	Главные опорные слова	Код ГОС	Второстепенные опорные слова	Код ВОС
1	2	3	4	5
Субъекты	*Ақсақ Ақан*	C01	*жігіт*	S01/01
	Кудря	C02	*жолдас*	S01/02
	Хамит	C03	*ел (қазақ елі)*	S02
	–	–	*Сейсенбаев*	S03/01
	–	–	*нұсқаушы*	S03/02
Объекты	–	–	*банды*	G01
	–	–	*ет*	G02
	–	–	*сөз*	G03
Место	*ауыл*	M01	*үй*	P01
	–	–	*поселке*	P02
	–	–	*Шиелі*	P03
	–	–	*Ақмола*	P04
	–	–	*Политбюро*	P05
	–	–	*арасы*	P06/01
	–	–	*қасы*	P01/02
	–	–	*жағы (оң, сол)*	P01/03
Время	–	–	–	–

Таблица 29. **Таблица основного статического содержания (ТОСС по PUB-06)**

Конкретное значение опорных слов	Группа опорных слов			
	Главные опорные слова	Код ГОС	Второстепенные опорные слова	Код ВОС
1	2	3	4	5
Субъекты	*оқушы*	C11	*пән*	S11
	–	–	*мұғалім (пән мұғалімі)*	S12
Объекты	*тіл (ана тілі)*	K11	*тақырып (еркін тақырып)*	G11
	әдебиет (тіл мен әдебиет)	K12	*мүше (секция мүшесі)*	G12
	апта (апта күндері)	K13	*қазына (асыл қазына)*	G13
	апталық (пән апталығы)	K14	*сабақ (ашық сабақ)*	G14
	пән (пән апталығы)	K15	*білім (білім тереңдету)*	G15
Место	*мектеп (орта мектеп)*	M11	*класс (7[а] - класында)*	P11
	–	–	*Зал (акт залында)*	P12
Время	*күн*	T11	–	–
	(алғашқы күн)	T11/01	–	–
	(екінші күн)	T11/02	–	–
	(үшінщі күн)	T11/03	–	–
	(төртінші күн)	T11/04	–	–
	(бесінші күн)	T11/05	–	–
	(сенбі күні)	T11/06	–	–
	кеш (қорытынды кеш)	T12	–	–

Таблица 30. **Таблица основного статического содержания (ТОСС по NAU-15)**

Конкретное значение опорных слов	Группа опорных слов			
	Главные опорные слова (ГОС)	Коды	Второстепенные опорные слова (ВОС)	Коды
Субъекты	–	–	–	–
Объекты	*кристалл*	K21	*әдемілік*	G21
	минерал	K22	*әсер*	G22
	ерітінді	K23	*көз*	G23
	газ	K24	*құрам*	G24
	зат	K25	*қысым*	G25
	қатты зат	K25/01	*жылдамдық*	G26
	балқыған зат	K25/02	*молекула*	G27
	бөлшек	K26	–	–
	тур	K27	–	–
Место	–	–	*табиғат*	P21
	–	–	*су*	P22
Время	–	–	–	–

Существует огромное число исследований по проблемам представления знаний в системах, работающих с естественным языком [86, 158, 197]. Нашим задачам более всего подходит "ролевой" подход, т.е. дифференцирование слов по той роли, которую выполняют их денотаты в реальных ситуациях.

Если считать, что в микроситуации можно выделять некоторые далее неделимые составляющие, которые мы решили называть элементарными составляющими микроситуации (ЭСМС), то, анализируя тексты различного рода, можно выделить такие ЭСМС, как "действие", "состояние", "место", "время", "средство", "причина", "цель", "характеристика", "детализация". Каждой такой ЭСМС в абзаце могут соответствовать слово или сочетания слов (глаголы, причастия, существительные и их сочетания). Тогда специфика каждой микроситуации будет задаваться определенной последовательностью ЭСМС и индивидуальным наполнением каждой из них соответствующими денотатами. В процессе изучения различных текстов выделяются возможные типы абзацев и каждому абзацу приписываются в определенном порядке группы слов, т.е. слова-конкретизаторы, замещающие в абзаце соответствующие денотаты микроситуаций.

Слова, отражающие ЭСМС, в определенной степени достраивают "костяк" содержания текста, заданный таблицей основного содержания, но они не завершают его полное построение. Это объясняется тем, что, кроме опорных слов и слов-конкретизаторов, в любом тексте существует еще одна группа слов, которую называют словами-наполнителями. Они "наполняют" полученное содержание качественными признаками, т.е. указывают качества предметов, названных именами существительными, признаки действий, определяемые в тексте глаголами и причастиями [126, с.248250].

Экспериментальные данные, представляющие собой слова-конкретизаторы и слова-наполнители для казахских текстов, в этой работе не рассматриваются. Отметим только то, что исследования по этому вопросу для различных функциональных стилей казахского языка требует все еще тщательного изучения.

5.6. Средства связи в предложении, между предложениями в абзаце и между абзацами

Виды и средства синтаксической связи в словосочетаниях, а также в простых и сложных предложениях казахского языка достаточно хорошо изучены и освещены в академических грамматиках казахского языка. Кроме того, по этому вопросу защищены и кандидатские диссертации. Необходимо отметить исследование, проведенное Ж. Т. Кайшыгуловой "Қазақ мәтінінің күрделі фразалық тұтасым // абзац деңгейінде мүшеленуі" (руковод. Н. М. Уалиев) [140]. В этой работе перечислены средства связи отдельных предложений, которые создают целостность текста: *"Мәтін құрылымына кіретін сөйлемдердің байланыстылығын білдіретін формальды құралдар, қайталанатын сөздер, жіктеу (III жақ) және сілтеу есімдіктері, синоним, антонимдер, түбірлес сөздер, жалғаулықты шылаулар, демеуліктер т.б."* [140, с.11]. Эти средства связи могут также встречаться внутри простых и сложных предложений.

Учитывая суждения, имеющиеся в литературе по этой проблеме казахского языка, к формализуемым средствам связи предложений письменной речи нами отнесены следующие: 1) связи через формы слов; 2) связи с помощью различного рода союзов и союзных слов *(шылау сөздер);* 3) связи через порядок слов в предложении.

Из перечисленных средств связи мы остановимся только на тех, которые осуществляются с помощью различных союзов и союзных слов *(шылау сөздер).* Рассматривая такой вид связи, мы составили *табл. 31,* где отражены возможные перечни союзов и союзных слов казахского языка. В *табл. 31* указаны специальные коды, принятые нами условно, которые будут использованы в дальнейшем в процессе записи формул предложений. Задание каждого союза или союзного слова отдельным кодом связано с тем, что практически невозможно найти правила, которые бы определяли их выбор в предложении. В то же время, они являются важными элементами предложения, вносящими в смысл предложения дополнительные логико-смысловые, модальные или экспрессивные оттенки. Их выбор определяется субъективными особенностями автора текста.

На основании данных, полученных в результате анализа средств связи трех видов рассмотренных нами текстов, и, учитывая некоторые идеи, высказанные в работах [42, 55], мы можем предложить следующую классификацию формальных средств связи между предложениями в абзаце:

1) соединительные элементы;
2) повторение лексических форм;
3) замещения;
4) ассоциативные синонимические связи;
5) ассоциативные антонимические связи;
6) грамматические средства связи.

Не будем подробно останавливаться на анализе всех перечисленных типов связи, так как абзац для нас является минимальной смысловой единицей текста, в которой связи уже особым образом определены и заданы. Рассмотрим лишь те связи, которые помогут в дальнейшем определить формальным способом некоторые особенности самих абзацев.

Перечислим случаи, которые чаще других являются соединительными элементами связи между предложениями в абзаце.

1. Связи с помощью союзов и союзных слов: *және, оның үстіне, тағы да, неге десең, сонымен бірге, оның есесіне, оның орнына , солай бола тұрса да, әрине, шынын айтқанда, сөйтіп, сонымен, сондықтан, сол себепті, сол үшін, соның нәтижесінде;* разделительные союзы: *я, не, немесе, не болмаса, біресе, яки,* условный союз *егер* и т. д.

2. Связи с помощью слов, которые выступают в роли союзов – слова-местоимения: *қандай, сондай, қанша, сонша, кім, сол, не, қайсы, қайда, қайдан, содан, қалай* и т. д.

3. Связи с помощью различного рода связок – логико-смысловых скрепов:

а) результат, итог: *қорытсақ, қорыта айтқанда, қорытындысы, нәтижесінде, ақырында, соңында, соң, сосын, айта келе, айта берсек, дәл осы, тек қана, осының барлығы* и т. д.;

б) сопутствующие обстоятельства, дополнения: *оның үстіне, одан басқа, ол аз болса, сонымен бірге, онан әрі, әрі қарай, ары, мұның үстіне, онан кейін, одан әрі, одан басқа, осыдан басқа, бола тұра, солай бола тұра, соған ұқсас, не үшін, сол үшін, дәл солай* и т.д. ;

в) возражения: *бірақ, бірақ та, алайда, дегенмен, солай бола тұрса да, әйтсе де, әйтпесе де, сөйтсе де, әйткенмен, керісінше, өйткенмен, сонда да* и т. д. ;

г) оценки достоверности сообщения, подтверждения, положительной аргументации: *дұрысында, шындығында, дұрысы, шыны, шынында, расында, асыра айтса, кем дегенде, сүйте тұра, солай бола тұра, табиғи, әрине, орынды, өте орынды, мәні мынада, одан да артық* и т. д.;

Таблица 31. **Союзы и союзные слова, являющиеся средствами связи слов в предложении (и между предложениями)**

Союзы и союзные слова	Код	Союзы и союзные слова	Код	Союзы и союзные слова	Код
1	2	1	2	1	2
я	Ш101	*еке*	Ш131	*бенен*	Ш161
ал	Ш102	*кел*	Ш132	*бетер*	Ш162
ба	Ш103	*кет*	Ш133	*бұрын*	Ш163
бе	Ш104	*кет*	Ш134	*бірақ*	Ш164
да	Ш105	*қай*	Ш135	*бірге*	Ш165
де	Ш106	*құй*	Ш136	*бірде*	Ш166
ма	Ш107	*мен*	Ш137	*ғұрлы*	Ш167
ме	Ш108	*пен*	Ш138	*дейін*	Ш168
на	Ш109	*соң*	Ш139	*демек*	Ш169
не	Ш110	*шық*	Ш140	*жуық*	Ш170
па	Ш111	*іші*	Ш141	*кейде*	Ш171
пе	Ш112	*асты*	Ш142	*кейін*	Ш172
та	Ш113	*әлде*	Ш143	*қатар*	Ш173
те	Ш114	*басы*	Ш144	*қарай*	Ш174
ша	Ш115	*бері*	Ш145	*құрлы*	Ш175
ше	Ш116	*гөрі*	Ш146	*мейлі*	Ш176
-ау	Ш117	*ғана*	Ш147	*менен*	Ш177
-ай	Ш118	*егер*	Ш148	*өйтсе*	Ш178
-ақ	Ш119	*екеш*	Ш149	*пенен*	Ш179

Окончание табл. 31

1	2	1	2	1	2
-ды	Ш120	*жаны*	Ш150	*сайын*	Ш180
-ді	Ш121	*және*	Ш151	*салым*	Ш181
-ты	Ш122	*қана*	Ш152	*сияқты*	Ш182
-ті	Ш123	*қасы*	Ш153	*сүйтсе*	Ш183
әлі	Ш124	*олай*	Ш154	*сымақ*	Ш184
әрі	Ш125	*тағы*	Ш155	*таман*	Ш185
бен	Ш126	*талу*	Ш156	*тарта*	Ш186
бер	Ш127	*үсті*	Ш157	*түгіл*	Ш187
бол	Ш128	*үшін*	Ш158	*шақты*	Ш188
гой	Ш129	*яғни*	Ш159	*шейін*	Ш189
еді	Ш130	*әйтсе*	Ш160	*алайда*	Ш190
арқылы	Ш191	*секілді*	Ш206	*ал ендеше*	Ш221
біресе	Ш192	*сықылды*	Ш207	*бірақ сондада*	Ш222
ғұрлым	Ш193	*тәрізді*	Ш208	*егер де*	Ш223
ендеше	Ш194	*әйткенде*	Ш209	*ең болмаса*	Ш224
қаралы	Ш195	*әйткенмен*	Ш210	*не болмаса*	Ш225
немесе	Ш196	*әйтпегенде*	Ш211	*не себеп*	Ш226
себебі	Ш197	*дегенмен*	Ш212	*олай болса*	Ш227
туралы	Ш198	*өйткенмен*	Ш213	*оның үстіне*	Ш228
тұрмақ	Ш199	*сондықтан*	Ш214	*оның есесіне*	Ш229
тұрсын	Ш200	*алда-жалда*	Ш215	*оның орнына*	Ш230
шамалы	Ш201	*әйтсе де*	Ш216	*сонда да*	Ш231
әйтпесе	Ш202	*олай да*	Ш217	*сол себепті*	Ш232
бойымен	Ш203	*өйтсе де*	Ш218	*сонымен бірге*	Ш233
болмаса	Ш204	*сүйтсе де*	Ш219	*солай болса да*	Ш234
өйткені	Ш205	*тағы да*	Ш220	*солай бола тұра* и т.д.	Ш235

д) причинно-следственные связи: *туындайды, сондықтан, демек, солай болғасын, олай болса, солай болғандықтан, солай болған соң, сонда, әсерінен, нәтижесінде, оның орнына, солай бола тұра, осы себептен, дәл сол себептен* и т. д.;

е) обстоятельственные отношения: *бұл жерде, мұнда, мында, содан бері, осыған дейін, онда, енді, осында, сонымен бірге, сонда да, барлығына тән, барлығына ортақ* и т. д.;

ж) субъективное отношение автора к высказанному в тексте: *тән, жиі, сирек, оңай, жеңіл, қиын, көруге оңай, айтуға оңай, жеңіл мінез, байқауға қиын, шындық, айта кету қажет, күмән тудырады, күмәнсыз* и т. п.

К группе логико-смысловых скрепов можно отнести также следующие:

з) перифразы: *басқа сөзбен (айтсақ), басқаша айтсақ, өзге сөзбен (айтқанда), шындығында, шынайы айтса, сөз болып отырған* и др.;

и) перечислительные конструкции: *біріншіден, екіншіден, бір жағынан (алғанда, қарастырғанда), басқа жағынан (алсақ, келсек), бірінші, екінші, үшінші* и т.д.;

к) слова, предшествующие приведению примера, факта или некоторой аргументации: *соның ішінде, сонымен қатар, сонымен бірге, сондай-ақ, мысал ретінде, мысалы, сонымен, "сонымен, мысал"* и т.д.

Среди таких средств связи, как "повторение лексических форм", могут участвовать повторение слов и словосочетаний с отсылкой, например *сол тәрізді, сол сияқты, берілгендей, қарастырылғандай, (жоғарыдағыдай), бұл, одан басқа, одан бөтен, тіпті, басқа, қалғандары* и т. д.

Все эти и другие соединительные элементы должны быть специально заменены условно принятыми кодами для программы порождения казахского текста.

Рассматривая формальные средства связи между абзацами в пределах некоторого текста, многие исследователи отмечают, что такие средства в целом совпадают с теми формальными средствами, которые соединяют предложения в абзаце [162, с.182; 184, с.37]. Здесь достаточно отметить, что частое употребление тех или иных средств связи зависит, в основном, от типа текстов (повествование, описание, рассуждение).

Для рассматриваемых нами типов текстов весьма характерными средствами межабзацных связей являются ассоциативные синонимичные связи, которые могут иметь следующие разновидности:

1. Связь через общеупотребительные или словарные синонимы: *ағайын – туыс – жақын, адам – кісі, айқын – анық – нақты, айырмашылық – өзгешелік – ерекшелік, аласа – тапал – қысқа – кішкене – мықыр – қортық, аман – есен – сау – саулық – сәлемет, ас – тамақ – дәм – тағам, әдіс – тәсіл – амал – жол* и т. д.

2. Связь через различного рода контекстуальные синонимы:

а) через понятия, объединяемые ассоциативными синонимами типа "род" – "вид": *колхозшы – малшы, егінші, ғалым – тілші, әдебиетші, тарихшы ..., ... ғылым – тіл, математика, медицина, философия, ..., дәрігер – хирург, стоматолог, терапевт, ...* и т.д.;

б) связь через ассоциативные слова типа "целое" – "часть" и "часть" – "целое": *адам – қол; жапырақ – өсімдік; үй – бөлме, терезе, есік; ағаш – табиғат* и т. д.;

в) связь через ассоциативные слова, объединенные по принципу разной степени одного и того же признака или интенсивности обозначения действия: *жылы – ыстық – күйіп тұр, жанып тұр – суық – салқын; баяу – тез – жылдам – шапшаң – жүйрік (жүріс); сұйық – қою – қатты (зат)* и т. д.;

г) связь через ассоциативные слова, объединенные по принципу ассоциативного переноса понятии по таким категориям, как "причина" – "следствие", "предмет" – "материал" – "качество", "действие" – "объект" – "деятель": *суық – тоңу (мұздау, қату); ыстық – еру – терлеу; шелек – темір – ыдыс; үй – ағаш – тас – шым – көркем – сұлу – биік; ату – мылтық; ұшу – құс – ұшақ – ұшқыш* и т. д.;

д) связь через ассоциативные слова, объединенные любой другой общностью коннотативных значений: *ауру – дәрігер – аурухана, қасқыр – ұлу, ит – ұру, құс – ұшу* и т. д.

Ассоциативные синонимические связи между абзацами являются основой статического содержания текста и формируют главные и второстепенные опорные слова таблицы основного содержания текста (см. *табл. 27–30).*

Существенное отличие связей между предложениями в абзаце от связей абзацев в тексте заключается в том, что первые, как правило, фиксируются в соседних

предложениях абзаца [55, с.17-18], в то время как абзацы могут быть связаны дистантными связями [184,с.38–40].

Мы принимаем случай, когда формальные средства связи между абзацами в тексте совпадают с такими же средствами для связи предложений в абзаце.

5.7. О семантико-синтаксическом языке для записи формул предложения, абзаца и текста

Абзац – это некоторый синтаксический шаблон, фиксирующий логику мысли автора при выражении отношений между некоторыми составляющими фрагмента психической ситуации, отраженного в абзаце. Поэтому, изучая проблему организации текстов из элементов разных уровней, мы должны найти способ фиксации в абзацах реальных текстов всех его компенентов. Возникает необходимость создания исскусственного языка, который мог бы служить некоторой моделью естественного языка (ЕЯ).

Известны три основных подхода к такому моделированию. Мы остановимся только на том случае, когда моделирование ЕЯ сводится к статистическому выбору из текстов наиболее употребительной лексики и грамматических правил, относящихся к узкой предметной области.

В отличие от существующих искусственных языков, основанных на базе ЕЯ, предлагаемая версия должна иметь средства экспликации различного рода повторов, опущенных частей смысла, средств связи между предложениями и абзацами и ряд других возможностей [173, с.21–29].

Вкратце остановимся на структуре и составе предлагаемого семантико-синтаксического языка (СЕМСИНТ). Мы будем придерживаться общих положений, разработанных А. В. Зубовым в книге "Проблемы порождения текста" [127; 128, с.26–38].

В нашем представлении такой искусственный язык (ИЯ) должен иметь возможность описать абзац текста как некоторую семантико-синтаксическую единицу, учитывающую статический и динамический аспекты, т.е. ИЯ может строить некоторую формулу абзаца. Если любой абзац состоит из цепочки взаимосвязанных предложений, то искомый язык должен содержать правила, позволяющие учитывать семантические и синтаксические отношения между членами предложения, т.е. он должен строить семантико-синтаксическую формулу предложения. В нем также должны участвовать элементы, отражающие логико-смысловые связи между предложениями одного абзаца и между абзацами. И наконец, для описания содержания всего текста через содержание отдельных абзацев в ИЯ должны быть средства для создания из формул абзацев некоторой формулы текста.

Следуя положениям А. В. Зубова [128, с.26-38], мы для описания содержания текста в ИЯ СЕМСИНТ также включили следующие составляющие:

1) алфавит языка СЕМСИНТ;

2) набор элементов для записи семантических отношений между членами предложения;

3) набор элементов для записи синтаксических отношений между членами предложений;

4) набор элементов для фиксации логико-семантических отношений между предложениями одного абзаца;

5) набор элементов для связи в абзаце статической и динамической составляющих текста;

6) семантико-синтаксическую формулу абзаца (СЕСФА);

7) набор элементов для фиксаций связей между абзацами одного текста;

8) семантико-синтаксическую формулу текста (СЕСФТА).

Точка зрения о том, что предложение – это знак, имеющий план выражения (линейная цепочка имен) и план содержания (смысл), поддерживается многими учеными. Принимая эту точку зрения, следует выяснить, как смысл предложения образуется из значений входящих в него слов, с одной стороны, и как этот смысл проявляется, выражается конкретными синтаксическими схемами – с другой.

Не вдаваясь в подробности этого вопроса, отметим, что наиболее распространена теория, трактующая процесс порождения предложения от семантики к синтаксису. Одним из представителей этой теории явился Ч. Филлмор. В своих работах семантическую структуру предложения он называет глубинной структурой [291, 292]. Наиболее близка этому направлению теория В. Г. Гака. Глубинную структуру он определяет как лексико-синтаксическую структуру, изоморфную ситуацию и называет прямой номинацией ситуации, а поверхностную структуру – косвенной номинацией [73].

Процесс перехода от семантической структуры к синтаксической сильно отличается в разных моделях и затрудняет их разъяснение, но общим для всех моделей является то, что процесс построения предложения начинается в них с организации его синтаксической структуры. Одной из таких наиболее распространенных моделей является модель Н. Хомского [300]. Согласно этой теории, предложение проходит в процессе порождения три этапа: категориальный, субкатегориальный и лексемный.

Анализируя приведенные подходы к проблеме организации смысла предложения, следует отметить их общие недостатки: они опираются почти полностью на автосемантичные предложения, т.е. на предложения, имеющие смысл вне текста. Кроме этого, большинство рассмотренных теорий касается вопроса организации лишь простых предложений.

В нашем исследовании предложение рассматривается как несамостоятельная часть абзаца, которая отражает какую-то часть фрагмента ситуации, представленной в описании абзаца. Необходимо предположить, что за каждым словом предложения стоит некоторый предмет, факт, явление объективной действительности. Кроме того, мы не считаем, что "глагол" является основным элементом семантико-синтаксической структуры предложения. В нашем случае, это опорные слова – существительные, к которым "подбираются" уже глаголы. Они и есть основное семантико-синтаксическое содержание абзаца и каждого его предложения.

В целом, предлагаемая модель является развитием комбинаторного подхода, учитывающего семантический подход Ч. Филлмора в его модернизированном виде [291], теорий Ю. Д. Апресяна и В. В. Богданова [12, 43], а также синтаксических идей В.М.Солнцева [261].

Семантическая формула предложения, как считают многие исследователи, должна отражать семантико-синтаксические роли конкретных участников реальной (или воображаемой) ситуации, отоброжаемой предложением. Такие участники ситуации в тексте задаются опорными и иными словами и именными словосочетаниями.

При анализе предложения в абзацах исследуемых текстов можно выделить семантические роли участников ситуации (или семантические функции), отображенной в предложении. Кроме того, для написания семантико-синтаксической формулы предложения нам необходимы специальные коды, по значениям которых узнаются содержание каждой семантической функции, а также их синтаксические значения в СЕМСИНТ (*табл. 33*). В целях увеличения точности отражения формулой элементов ситуации можно ввести условные обозначения семантических подклассов знаменительных классов слов, конкретные значения которых даны в таблице "Семантические классы знаменительных классов слов казахского языка" (см. *табл. 3 по–10)*. Коды, соответствующие названиям частей речи (их скращения) и семантическим подклассам слов казахского языка, приведены в *табл. 32.*

Выделяемые нами роли участников ситуации будут относиться к группам слов, включающим имена существительные (именные словосочетания) и все относящиеся к ним определители, выраженные именами прилагательными, причастями, порядковыми числительными и местоимениями. В дальнейшем такую группу слов мы будем называть **аргументной группой.** В формуле предложения она должна состоять из цепочки кодов семантических подклассов определителей и существительных, соединяемых знаками * (звездочка), ограниченных слева знаком < (меньше) и справа знаком > (больше).

Для обозначения места глагола-сказуемого в формуле предложения можно использовать какой-либо код, например R, за которым необходимо поставить цифру, указывающую семантическую валентность глагола. Уточнение типа глагола осуществляется путем указания вслед за кодом ЕТ семантического подкласса через индексы ijk (ijk=101, 102, 103, ..., 999), которые указывают на номер семантического подкласса (см. *табл. 4*).

Таблица 32. **Коды семантического подкласса слов казахского языка, используемые в языке СЕМСИНТ**

№ п/п	Классы слов	Код семан. подкл.	№ п/п	Классы слов	Код семан. подкл.
1	Существительное	$ЗТ_{ijk}$	5	Причастие	$ЕШ_{ijk}$
2	Глагол	$ЕТ_{ijk}$	6	Деепричастие	$КШ_{ijk}$
3	Прилагательное	$СН_{ijk}$	7	Местоимение	$ЕС_{ijk}$
4	Наречие	$УС_{ijk}$	8	Числительное	$СА_{ijk}$

Заметим, что в формуле предложения наречия будут представляться в виде третьего составного элемента после аргументной группы и глагола, а все входящие в формулу элементы, т.е. семантические функции, глагол и наречия, соединяются между собой знаком + (плюс) в строгом соответствии с порядком следования слов в предложении.

Учитывая результаты исследования Ч. Филлмора, Ю. Ю. Апресяна, В. В. Богданова, а также анализа трех разновидностей казахских текстов, мы выделили семантические роли (семантические функции) участников ситуации (*табл. 33).*

Таблица 33. **Семантические функции, используемые в языке СЕМСИНТ и их синтаксические значения**

№ п/п	Семантическая функция	Содержание семантической функции. Характеризует аргументную группу как отражение в предложении в следующих событий:	Код
1	2	3	4
1	Субъект	Активного одушевленного инициатора события (действия, состояния, процесса) вместе с его признаками, выражаемого глаголом предложения	AAG
2	Предмет	Активного неодушевленного инициатора события – предмета вместе с его признаками	AH1
3	Понятие (общее)	Активного неодушевленного инициатора события – общего понятия вместе с его признаками	AS2
4	Понятие (научное)	Активного неодушевленного события – научного понятия вместе с его признаками	AS1
5	Процесс	Активного неодушевленного инициатора события – процесса с его признаками	AS3
6	Деятель (неодушевленный)	Активного неодушевленного инициатора события – природного или стихийного явления или предмета вместе с их признаками	AEL
7	Объект	Одушевленного объекта некоторого события вместе с его признаками	AO
8	Получатель	Одушевленного адрестата, получателя некоторого события, выражаемого глаголом предложения	AB1
9	Адресат	Неодушевленного адрестата некоторого события	AB2
10	Объект	Неодушевленного объекта некоторого события, выражаемого глаголом предложения	AH
11	Место	Места совершения некоторого события, выражаемого глаголом предложения	ALK
12	Средство	Средства выполнения события, выражаемого глаголом предложения	AMD
13	Инструмент	Орудия или инструмента выполнения события, выраженного глаголом предложения	AIN
14	Состав	Состава или содержимого какого-либо предмета, выражаемого в предложении другой аргументной группой	AKM
15	Определитель	Некоторого свойства какого-либо предмета, выраженного в предложении своей аргументной группой	AD1
16	Время	Времени осуществления события, выражаемого глаголом предложения	ATM
17	Способ	Способа реализации события, выраженного глаголом предложения	AAD
18	Условие	Условия осуществления события, выраженного глаголом предложения	AIF
19	Причина	Причины осуществления события, выраженного глаголом предложения	ACS
20	Цель	Цели выполнения события, выраженного глаголом предложения	AAM
21	Вставка	Некоторой вспомогательной информации, разъясняющей или уточняющей роли, выполняемые группами предложений (примеры, таблицы, графики, рисунки)	IN_i (i=1, 2, 3, 4)

Искусственный язык, принятый на основе ЕЯ, разработанный А. В. Зубовым и названный семантико-синтаксическим языком, вполне может удовлетворять требованием, которые ставятся при формализованной записи содержания казахского текста.

По А. В. Зубову, алфавит языка СЕМСИНТ представляется в следующем виде:

1) все русские буквы алфавита (кроме ё);

2) все латинские буквы;

3) десятичные цифры: 0, 1, 2, ..., 9;

4) орфографические знаки: .(точка), :(двоеточие), ?(знак вопроса), ‘(апостроф), “,”(кавычки), (,)(левая и правая круглые скобки), [,](левая и правая квадратные скобки), ...(многоточие), – (черточка);

5) знаки арифметических действий: +(сложение), /(деление), · (умножение), – (вычитание), · · (возведение в степень);

6) знаки логических действий: ¬ (отрицание), & (конъюнкция), |(дизъюнкция), > (больше), < (меньше), = (равно).

Некоторые формальные средства связи, которые приняты нами для записи формулы предложения таковы:

а) синтаксическая связь слов в субстантивно-субстантивных словосочетаниях в формуле-предложении обозначается знаком наклонной черты (/);

б) однородные члены предложения, так же, как и аргументные группы, в формуле-предложении соединяются логическим знаком & (конъюнкция);

в) знаком | (вертикальная черта) обозначаются союзы *не, немесе, не болмаса,* когда они употребляются внутри аргументной группы или соединяют две (или более) аргументные группы;

г) знак = (равно), стоящий после слова, показывает, что слово до знака “равно” восстановлено из предыдущего контекста и при наполнении формулы конкретной лексикой это слов может быть опущено.

Итак, приведенные сведения об алфавите языка СЕМСИНТ, а также некоторые формальные средства связи слов в предложении и между предложениями абзаца будут применяться в дальнейших наших разработках по этой проблеме.

Семантико-синтаксическая формула предложения (СЕСФП), которая может быть представлена в виде C_i (“смысл”, где i=1,2,3,...) строится с помощью 21 семантической функции (*табл. 33*), разделенной по значению требуемых ими поверхностных падежей, а также кодов союзов, союзных слов (*табл. 31*) и других положений. В формуле предложения будут использованы коды семантических подклассов казахских существительных, глаголов, прилагательных, причастий, деепричастий, наречий и местоимений (см. *табл. 3–10).*

На основании формул предложений (формулу предложения обозначим через C_i, где i=1,2,3, ...) строится уже семантико-синтаксическая формула абзаца (СЕСФА) как детерминированная последовательность формул предложений C_i(СЕСФП). Если СЕСФА представить в виде M_i (“микротема”), то имеем:

$$M_i = C_1 \,\&\, C_2 \,\&\, C_3 \,\&\, \ldots \,\&\, C_n.$$

Следует полагать, что у каждого автора, пишущего в той или иной узкой области знания, существует ограниченное число таких последовательностей для выражения некоторых типичных фрагментов ситуации.

Известно, что порядок следования абзацев зависит от типа текста, намерений автора, цели создания текста, а также от получателя текста и т.п. Между абзацами

текста должны существовать как детерминированные правила, так и вероятностные факторы. С учетом изложенного и того, что текст состоит из ограниченного количества абзацев, семантико-синтаксическую формулу текста (СЕСФТ) в свернутом виде можно представить как последовательность СЕСФА:

$$Ti=M_1(D_1/B_1)\&M_2(D_2/B_2)\&M_3(D_3/B_3) \ldots \&M_n(D_n/B_n),$$

где Ti – тема всего текста; Mi – микротемы, заключенные в абзацах i; Di – детерминированные правила, связывающие абзац "i" с абзацем "(i-1)"; Bi – вероятностые факторы, связывающие абзац "i" с абзацем "(i-1)".

Практическое написание формул предложений, абзаца и конкретного казахского текста вполне возможно, если воспользоваться материалами нашего исследования. А для их реализации на компьютере необходимы дополнительные исследования специалистов по инженерной лигвистике.

В заключение предлагаем принципиальный алгоритм (укрепненную блок-схему) перехода от естественного казахского текста к его формальному представлению, сопровождая его конкретным казахским предложением.

Покажем работу принципиального алгоритма на примере.

Блок А.

Пусть вводится в память компьютера следующий текст: *"Үлкен жүк тасыйтын машина кешкі сағат алтыда теміржол көпірінен өтті. Ол Алматыға азық-түлік жеткізеді".* (*"Большая грузовая машина в шесть часов вечера переехала железнодорожный мост. Она доставит продовольствие в Алматы"*).

Блок Б.

Читаем очередное предложение: *"Үлкен жүк тасыйтын машина кешкі сағат алтыда теміржол көпірінен өтті."*

Блок В.

Заменяем все слова этого предложения лексико-грамматической и семантической информацией из автоматического словаря казахского языка (АСКЯ). В словаре может быть указано больше параметров, чем количество формальных признаков в замене. В этом случае условимся указать только некоторые, например, принадлежность данного слова к определенному классу – существительное (ЗТ), глагол (ЕТ) и т.д., а также код семантического подкласса:

1) рассмотрим первое слово предложения: *үлкен* – прилагательное – (СН), качественное прилагательное – (СCijk) и по семантической классификации (*см. табл. 7)* относится к "прилагательным, означающим объем предмета", код которого "СС23", данное слово в именительном падеже, т.е. условный код падежа – (АС) от *"атау септік";* таким образом, в формуле предложения слово *үлкен* будет заменено:

(СН)(АС)(СС23).

Аналогично этому слову производим замену и для остальных слов предложения;

2) *жүк* – существительное (ЗТ), именительный падеж (АС), нарицательное (ЖЕ), неодушевленное (ЖЗ), семантический класс – предметы-артефакты, используемые человеком – ЗТ315 (см. *табл. 3)* и др.; вместо слова *"жүк":*

(ЗТ)(АС)(ЗТ315);

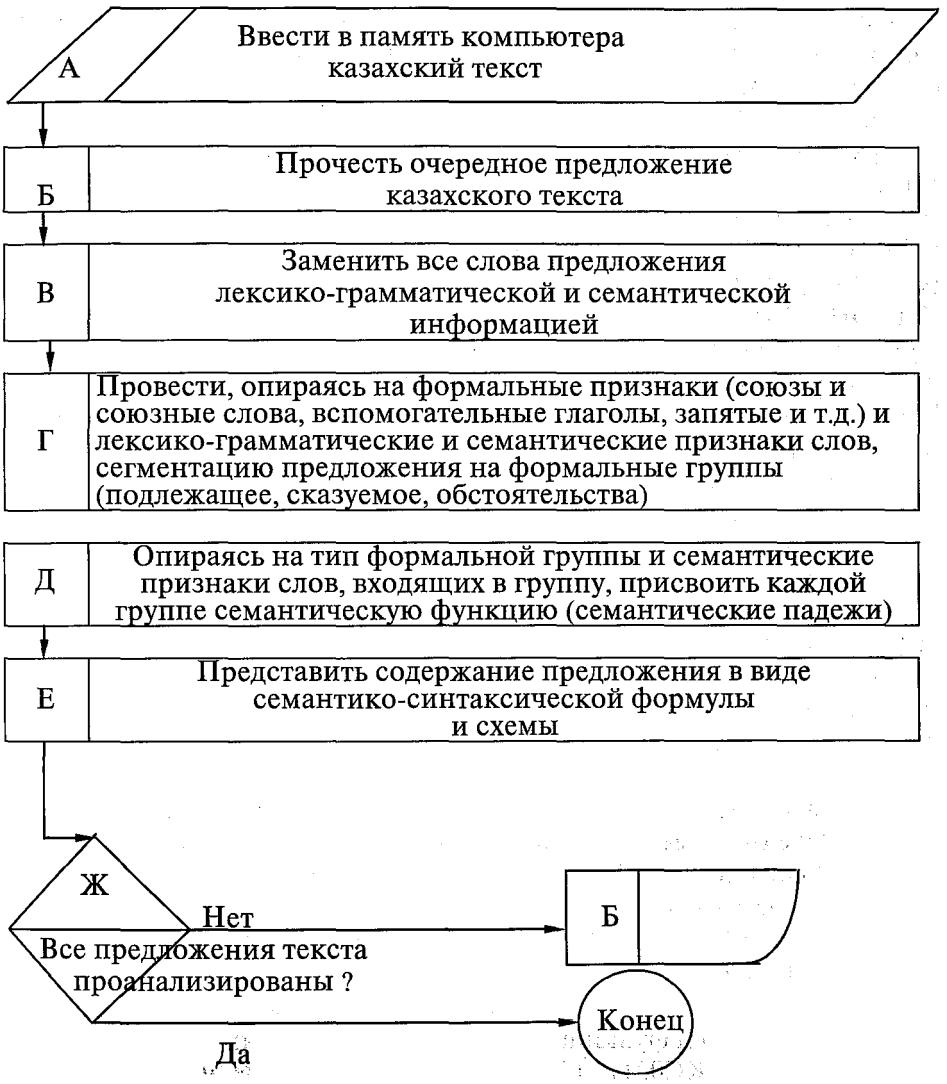
3) *тасыйтын* – глагол (ЕТ) в форме причастия, прошедшее время (ӨШ), по семантическому подклассу как причастие означает физическое действие, свойственое живому и неживому в природе – ЕШ081 (см. *табл. 5)* и др., т.е. *"тасыйтын"*

(ЕТ)(ӨШ)(ЕШ081);

4) *машина* – сушествительное (ЗТ), единственное число (АС), нарицательное (ЖЕ), неодушевленное (ЖЗ), семантический подкласс ЗТ308; согласно *табл. 3,* это слово относится к подклассу: объект, используемый для передвижения человеком; итак, слово "*машина*" – ЗТ)(АС)(ЗТ308);

5) *кешкі* – прилагательное (СН), именительный падеж (АС); семантический подкласс как относительного прилагательного, согласно *табл. 8* характеризует объект по его отношению ко времени (ҚС031). Таким образом слово "*кешкі*" – (СН)(АС)(ҚС031);

Блок-схема 2. **Принципиальный алгоритм перехода от естественного казахского текста к его формальному представлению**



6) *сағат* – существительное (ЗТ), именительный падеж (АС), нарицательное (ЖЕ), как существительное, означающее временное поняие по *табл. 3,* – (ЗТ601); само слово в кодах запишется:

(ЗТ)(АС)(ЗТ601);

7) *алтыда* – числительное (СА), местный падеж (ЖС), количественное *(есептік сан есім)* (ЕК), т.е. все слово – (СА)(ЖС)(ЕК);

8) *теміржол* – существительное (ЗТ), единственное число (АС), нарицательное (ЖЕ), по семантической классификации это слово можно отнести к подклассу наука: "понятия, связанные с устройством отдельного механического объекта" и его код ЗТ117 (см. *табл. 3);*

итак, слово "*теміржол*" – (ЗТ)(АС)(ЗТ117);

9) *көпірінен* – существительное (ЗТ), исходный падеж (ШС), по семантике как и в предыдущем пункте – ЗТ117, а само слово: (ЗТ)(ШС)(ЗТ117);

10) *өтті* – глагол (ЕТ), прошедшее время (ӨШ), по семантическому подклассу относится к глаголам со значением каузации движения, а именно к пункту: "обозначающие преодоление преграды в движении" – ЕТ941,

то слово – (ЕТ)(ӨШ)(ЕТ941).

Теперь, заменив каждое слово в этом предложении соответствующими кодами, будем иметь:

(СН)(АС)(СС23) (ЗТ)(АС)(ЗТ315) (ЕТ)(ӨШ)(ЕШ081)
(ЗТ)(АС)(ЗТ308) (СН)(АС)(ҚС031) (ЗТ)(АС)(ЗТ601)
(СА)(ЖС)(ЕК) (ЗТ)(АС)(ЗТ117) (ЗТ)(ШС)(ЗТ117)
(ЕТ)(ӨШ)(ЕТ941).

Принятые в этой записи наборы кодов абсолютно условные, которые зависят от того, как был составлен автоматический словарь казахского языка (АСКЯ).

Блок Г. Проведем сегментацию предложения на группы или на так называемые аргументные группы. Рассматриваемое предложение можно разбить на следующие четыре аргументные группы:

1) *үлкен жүк таситын машина* – именная группа;
2) *кешкі сағат алтыда* – обстоятельственная группа;
3) *теміржол көпірінен* – именная группа;
4) *өтті* – глагольная группа.

Блок Д. Зная тип группы (именная, глагольная, обстоятельственная) и падежи существительных, входящих в группы и их семантические характеристики, присваеваем этим группам наименования семантических функций (субъект, предмет, понятия и т.д.) согласно *табл. 33* :

1) именная группа: "*үлкен жүк таситын машина*" – имя существительное в именительном падеже, неодушевленное; значит по значению семантической функции – "Предмет" вместе с признаками: *үлкен* и *жүк тасыйтын;* по данным *табл. 33* код "Предмет" – АН1. Теперь запишем предмет АН1, отбирая индекс класса и семантический подкласс:

АН1<(СН)(АС)(СС23)*(ЗТ)(АС)(ЗТ315)*(ЕТ)(ӨШ)(ЕШ081)*
(ЗТ)(АС)(ЗТ308)> ;

2) обстоятельственная группа: "*кешкі сағат алтыда*", рассматриваем аналогично по *табл. 33* и узнаем код функции – АТМ, который равен:

АТМ<(СН)(АС)(ҚС031) * (ЗТ)(АС)(ЗТ601) * (СА)(ЖС)(ЕК)>;

3) именная группа: "*теміржол көпірінен*" – неодушевленный объект некоторого события, который имеет код АН, тогда:

АН<(ЗТ)(АС)(ЗТ117) * (ЗТ)(ШС)(ЗТ117)>;

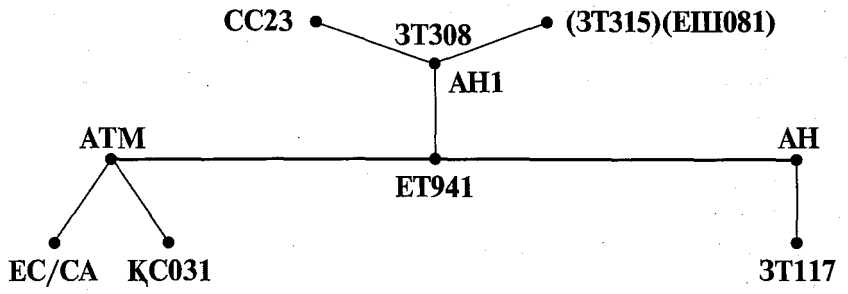
4) глагольная группа: *"өтті"* – (ЕТ)(ӨШ)(ЕТ941).

Блок Е.

В итоге все предложение записывается в виде семантико-синтаксической формулы, понятной компьютеру:

$С_1$=АН1<(СН)(АС)(СС23)*(ЗТ)(АС)(ЗТ315)
(ЕТ)(ӨШ)(ЕШ081) *(ЗТ)(АС)(ЗТ308)>+
+АТМ<(СН)(АС)(ҚС031)*(ЗТ)(АС)(ЗТ601)*
(СА)(ЖС)(ЕК)>+АН<(ЗТ)(АС)(ЗТ117)
(ЗТ)(ШС)(ЗТ117)>+ +(ЕТ)(ӨШ)(ЕТ941).

Представим теперь содержание предложения в виде следующей схемы:



Эту схему можно прочитать так: нечто (АН1) неодушевленное (ЗТ308), имеющее признак (СС23) и с определенным предназначением (ЕШ081), совершает физическое действие (ЕТ941) во время АТМ, связанное с некоторым объектом (АН), обладающим свойством (СН/ЗТ117).

Таким образом, можно было бы проанализировать и второе предложение, взятое для примера, и перейти от естественной его формы к формальному виду, так же, как и в случае с первым предложением.

Рассмотрев вкратце работу принципиального алгоритма, перехода от естественного казахского текста к его формальному представлению, можно отметить, что формализация содержания абзаца и всего текста осуществляется таким же образом. Для этого необходимо воспользоваться основными принципами формализации содержания казахского текста и другими данными, имеющими в предлагаемом исследовании, а для отдельных частных случаев, возможно, понадобятся дополнительные исследования.

ЗАКЛЮЧЕНИЕ

Поиски новых способов представления значения, смысла и содержания текстовых единиц в памяти лингвистических автоматов идут в самых разных направлениях [3, 191, 203, 228, 234]. Одно из них связано с последними достижениями науки о лингвистике текста, а именно с тем, что основной единицей письменной коммуникации является не предложение, а текст.

В этой связи, процессуальный, динамический подход к изучению языка, подход через изучение процессов, лежащих в основе порождения и понимания высказываний, позволит выявить ту роль, которую играют при формировании содержания текста лингвистические и энциклопедические, детерминированные правила и случайные факторы, языковые и логические структуры мышления [117, с.256; 154, с.51].

Ссылаясь на точки зрения лингвистов, психолингвистов и других специалистов на структуру текста, а именно на его составляющие, связь текста с языком, мыслью и реальной действительностью, можно выделить характерные черты текста, такие как наличие в нем единого содержания, определенных его единиц. Эти единицы текста связаны между собой как семантически, грамматически, так и логически. Причем, замечается неоднозначность таких связей между основными единицами данного текста.

Содержание текста следует понимать как отражение в сознании человека отношения линейной цепочки имен текста к трансформированной автором, в соответствии с замыслом, целями и условиями коммуникации, психической ситуации. Причем, релевантными для организации содержания текста являются две группы абзацев, выделенные по функционально-смысловому критерию и месту в структуре текста (начальные, медиальные и конечные).

Проанализировав высказывания ученых разных направлений и специальностей о роли вероятности в общей теории познания, процессах функционирования языка и организации различных текстов, мы можем сделать следующие выводы:

- все связи в реальном мире носят объективный характер, что проявляется в массе однородных явлений;
- язык, являясь саморегулирующей системой, при порождении речи функционирует, используя как детерминированные правила, так и случайные факторы;
- текст как результат речевой деятельности также содержит детерминированные и вероятностные составляющие;
- абзацы как "кванты" содержания и "застывшие" синтаксические шаблоны организуются в единое целое по вероятностно-алгоритмическому принципу.

Анализ конкретного материала, содержащего 102 минитекста в трех жанровых разновидностях позволил нам предложить формальный критерий – коэффициент важности $К_{важ}$, позволяющий отделить опорные слова казахского текста от остальных его слов. При этом удалось дифференцировать $К_{важ}$ на два критерия – $К^1_{важ}$ и $К^2_{важ}$, позволяющие выделить главные и второстепенные опорные слова в трех функционально-смысловых типах казахского текста. Следует отметить, что

значения коэффициентов этого критерия отличаются от соответствующих коэффициентов критерия важности, использованных для определения опорных слов в русскоязычных текстах [128, с.18-19].

Выделенные по указанному критерию опорные слова казахского текста (ГОС, ВОС), сыграли определенную роль в разработке построения "Таблицы основного статического содержания" в рассматриваемых нами 102 минитекстах.

Таким образом, подход к абзацу как к минимальной семантико-синтаксической единице текста позволил нам выделить в трех исследованных текстах определенные типы абзацев по их функционально-смысловому содержанию, структурному положению в тексте (начальный, медиальный, конечный), а также по предметно-логическому содержанию. Выделенные типы во всех указанных разновидностях текстов исследованы статистически.

Приступая к построению алгоритма порождения текста, разумно проявить некоторую осторожность и принять ряд ограничений [118, с.14]. Одним из таких ограничений исследования является выбор текстов конкретных авторов, представляющих три различных функционально-смысловых типа речи: рассуждение, повествование и описание. Следующее допущение, принятое нами, заключается в том, что тексты одного и того же автора могут быть представлены в виде ограниченного кванта содержания, в качестве которого мы приняли предметно-логическое содержание абзаца как основной единицы письменного текста. Анализ, с этой точки зрения, трех нами рассмотренных функционально-смысловых типов позволил выявить в них ограниченное число абзацев, из комбинации которых построены исследуемые тексты. Таким образом, установлено количество общих типов абзацев для всех исследованных текстов. В эту монографию не вошли данные, полученные на основе изучения вероятностей следования абзацев, которые также необходимы для составления алгоритмов порождения казахских текстов. На наш взгляд, такой вопрос требует особого рассмотрения для тюркоязычных текстов.

Важным моментом нашего исследования явилась семантическая классификация знаменательных классов слов казахского языка *(см. табл. 3–10)* и их кодификация, которая позволяет передавать некоторую жизненную ситуацию, описываемую в тексте.

Говоря об алгоритмах и программах порождения текста, нельзя не упомянуть известного ученого К. Шеннона, так как с его исследованиями связан алгоритм порождения цепочек слов, в котором учитываются вероятности употребления отдельных букв, пар букв, троек, четверок букв и т.д. [312]. Он показал, что чем на большую длину буквенных последовательностей учитывались связи, тем более осмысленными получались слова.

Достаточно большое число алгоритмов и программ было разработано исследователями для подтверждения тех или иных синтаксических и семантических моделей предложений. Но в них нет единого замысла, нет смысловых опорных точек, отсутствуют строгие смысловые критерии отбора синтаксических и лексических единиц [128].

Следует отметить, что формального построения правильных и осмысленных предложений можно достичь только при комбинировании детерминированных и случайных процессов порождения текста [129, с.182].

Интересны эксперименты, проведенные по составлению алгоритмов сочинения различных сказок на ЭВМ [72]. В основе всех попыток сочинения сказок

формализованным путем лежат исследования структуры сказок, проведенные В.Проппом [238]. В частности, в одном из алгоритмов порождения сказки в качестве исходной информации приняты следующие данные: перечень типов действующих лиц, перечень типовых характеристик персонажей, список возможных поступков каждого действующего лица, список возможных типов встреч и т. д. В итоге получается некоторая общая формула сказки [72, с.223]. Для более ясного представления об этом запишем небольшой отрывок из казахской сказки *"Әділ жаза" ("Справедливое наказание")* [165, с.73-79], основываясь на принципах этой формулы: *"Ерте-ерте ертеде, ешкі құйрығы келтеде, бір "А" болыпты. "А"-ның үш баласы бар екен. Ең үлкенінің аты – "Б", ортаншысының аты – "В", ал кенже баласының аты – "С" екен. "С" өте* <a>, <b>, *әрі* <g>, <r> *бала болып өседі. "А" ұйықтап жатып* <d> *көреді,* <d>*-да бір ғажайып* <e> *көреді.* <e> *күлсе, аузынан буда-буда* <g> *түседі, жыласа, көзінен* <z> *түседі. Міне, "А" осы* <e>*-ге ынтық болады. Дереу орнынан тұрысымен "Б" мен "В"-ны шақырып алып көрген* <d> *-ны айтып береді...".*

Участвующие в сказке условно принятые обозначения на самом деле означают: *"А" – патша* (царь), *"Б" – Асан* (имя старшего сына), *"В" – Усен* (имя среднего сына), *"С" – Хасан* (имя младшего сына), <a> – *сұлу* (красивый), <b> – *ақылды* (умный), <g> – *батыр* (смелый), <r> – *ақ көңілді* (веселый), <*d*> – *тұс* (сон царя), <*e*> – *құс* (волшебная птица), <*g*> – *гүл* (цветы), <z> – *моншақ-моншақ маржан* (бусы из драгоценных камней) и т. д.

Теперь *A, Б, B, C* и <a, b, g, r, *d, e, g,* z> и т. д., заменив другими действующими лицами вместе с их характеристиками, поступками и т. д., можно получить сколько угодно вариантов казахских сказок, созданных по этой формуле.

Стоит заметить, что общение с компьютером (ЭВМ) на естественном языке еще далеко от совершенства. В искусственных языках, понятных для машин, множество элементов формируется отношением эквивалентности, в естественном же языке, например казахском, такое объединение единиц проводится на основе толерантности или сходства.

Известно также, что основная особенность языковых множеств – наличие в нем нечетких, размытых границ, причиной которых является несовпадение структур естественного и искусственных языков. Такое несовпадение Р. Г. Пиотровский [228] считал следствием известных антиномий языка и речи, дискретности и непрерывности, синхронии и диахронии, языка и диалекта, предложенных еще Ф. де Соссюром [260].

Как мы полагаем, в использовании современной компьютерной техники в Республике Казахстан значительная роль должна отводиться созданию лингвистических процессоров, основанных на применении знаний, заложенных на естественном казахском языке. Благодаря этому, современные компьютеры не просто будут осуществлять накопление и поиск информации на национальном языке, но и будут понимать смысл запроса, а также выдавать результаты в виде текста и речи на казахском языке. Однако, создание таких лингвистических процессоров будет опираться на ограниченные языки определенных предметных областей (подъязыки).

Известно, что в каждом языке выделяются функциональные стили или "подъязыки", которые обладают, по сравнению с языком в целом, ограниченным количеством лингвистических единиц и определенной спецификой их употребления.

Выбирая наиболее частые и информационно насыщенные единицы языка, а также типовые контексты их употребления, мы можем построить такую модель языка (базовый язык), которая будет являться некоторым приближением к реальной системе подъязыка, порождающей тексты узкой тематики [231, с.21–23].

Особую ценность в республике в настоящее время представляет так называемая "деловая проза", написанная на казахском языке, а также армейские документы ("военная деловая проза") и национальный военный лексикон, который только начинает зарождаться в армейской жизни.

Таким образом, научные исследования в этом направлении, т.е. анализ и синтез текстов указанных документов, являются неотложной задачей ученых-казаховедов и специалистов по инженерной лингвистике.

Дальнейшее изучение проблемы порождения казахского текста требует более глубокого рассмотрения всех вопросов, касающихся семантико-синтаксических формул предложения, абзаца и текста в целом. И, конечно же, необходим практический анализ больших объемов текстов самых различных типов с помощью предложенного искусственного языка СЕМСИНТ для выявления и написания различных семантико-синтаксических формул этих текстов.

ПРИЛОЖЕНИЯ

Фрагмент 1

Алфавитно-частотный словарь слов по научно-популярному типу текстов[1]

№ п/п	Слово	Абсолют. частота F_1	Накоп. абсолют. частота f_1	Накоп. относит. частота. f_2	Абсол. частота текстов F
1	2	3	4	5	6
1	авиалайнер/зт	1	1	0,000040	1
2	Австралия/жа	1	2	0,000079	1
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
15	ағаш/зт	7	52	0,002059	4
16	ағыз/ет	9	61	0,002415	5
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
255	әбден/үс	1	2095	0,082947	1
256	әдемі/сн	1	2096	0,082987	1
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
300	баға/зт	2	2655	0,105119	2
301	бағала/ет	1	2656	0,105159	1
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
529	вагон/зт	2	5323	0,210753	2
530	вакуум/зт	1	5324	0,210793	1
- - -	- - - - - - - - -	- - -	-	- - - - - -	- - -
552	гектар/зт	12	5436	0,215227	3
553	гектардай/сн	2	5438	0,215307	2
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
594	ғажайып/сн	2	5551	0,219781	2
595	ғажап/сн	2	5553	0,219860	2
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
604	да/шл	155	5796	0,229481	39
605	дағдарыс/зт	1	5797	0,229521	1
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
2962	эрозия/зт	12	24947	0,987726	5
2965	Эстония/жа	2	24950	0,987845	1
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
2967	Юта/жа	1	24954	0,988003	1
2968	яғни/шл	26	24980	0,989033	15
- - -	- - - - - - - - -	- - -	- - -	- - - - - -	- - -
2970	ярус/зт	3	24986	0,989270	2

[1] Здесь показаны вид и структура словаря, где F_1 – абсолютная частота слов в 47 минитекстах, F – абсолютная частота минитекстов с данным словом. В конце слова после наклонной скобки записано условный индекс-сокращение: зт – *зат есім* (ат – *адам аты*, жа – *жер аты*), ет – *етістік*, шл – *шылау*, сн – *сын есім*, үс – *үстеу* и т.д.

Фрагмент 2

Распределительный алфавитно-частотный словарь слов по трем типам текстов[2]

№ п/п	Слово	Абсолютные частоты			
		Худож. F_1	Публ. F_2	Научн. F_3	Суммар. F
1	2	3	4	5	6
1	а/зт	0	4	0	4
2	а/од	0	2	0	2
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
38	ағаш/зт	14	4	7	25
39	ағыз/ет	1	0	9	10
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
49	адам/зт	60	22	26	108
50	адамгершілік/зт	0	13	0	13
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
719	әр/ес	12	49	74	135
720	әрбір/ес	1	24	6	31
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
803	байланыс/зт	3	12	38	53
804	байланысты/үс	0	57	66	123
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
1582	де/ет	215	153	101	469
1583	де/шл	126	89	81	296
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
2128	жас/сн	13	38	0	51
2129	жаса/ет	10	45	52	107
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
3649	қозы/зт	8	8	0	16
3650	қой/ет	43	42	20	105
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
4848	пайда/зт	5	2	52	59
4849	пайдалан/ет	0	10	73	83
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
6294	тығыз/сн	1	11	12	24
- - -	- - - - - - - - -	- - -	- - -	- - -	- - -
7239	ярус/зт	0	0	3	3

[2] О количественных данных каждого из типов можно узнать из книги.

Фрагмент 3

Алфавитно-частотный словарь слов-существительных по научно-популярному типу текстов[3]

№ п/п	Слово	F_1	F_2	№ п/п	Слово	F_1	F_2
1	2	3	4	1	2	3	4
Текст: NAU-01				Текст: NAU-18			
1	айырмашылық	4	2	1	адам	2	1
2	асбест	3	1	2	ангстрем	3	1
3	әдіс	3	1	3	ара	4	2
4	берилл	2	1	4	атом	14	3
5	ерекшелік	2	1	5	байланыс	2	1
6	жіп	2	1	6	ерекшілік	2	1
7	зат	5	2	7	жауап	2	2
8	кен	5	3	8	жыл	3	1
9	кристалл	10	3	9	зат	3	1
10	күй	2	2	10	кристалл	9	5
11	қасиет	6	3	11	кристаллохимия	2	1
12	құрылым	2	1	12	қасиет	3	2
13	минерал	11	4	13	қашықтық	3	2
14	тарау	2	1	14	құрылыс	3	1
15	түр	2	2	15	нәтиже	2	2
16	флотация	2	1	16	рентген	8	2
17	форма	2	1	-	- - - - -	-	-
Текст: NAU-07				Текст: NAU-22			
1	айырмашылық	1	1	1	алд	2	1
2	әдіс	8	2	2	анион	8	3
3	кен	2	1	3	ара	5	4
4	минерал	1	1	4	атом	9	3
5	руда	1	1	5	ауа	4	2
6	салмақ	1	1	6	әрекет	6	5
7	сыр	1	1	7	әсер	3	3
8	техника	1	1	8	байланыс	3	2
9	түр	1	1	-	- - - - -	-	-
Текст: NAU-09				Текст: NAU-30			
1	аппарат	3	1	1	аралық	4	3
2	бөлгіш	2	1	2	ауа	4	3
3	діңгек	2	1	3	әдіс	8	3
4	бөлгіш	2	1	4	бет	3	3
5	күш	3	1	5	депрессорлаушы	2	1
- -	- - - -	- -	- -	- -	- - - - - -	- -	- -

[3] F_1 – абсолютная частота слов в отдельном минитексте, F_2 – абсолютная частота абзацев в каждом из 47 минитекстов (число абзацев, где встретилось данное слово).

Фрагмент 4

Алфавитно-частотный словарь слов-существительных по художественному типу текстов[4]

№ п/п	Слово	F_1	F_2	№ п/п	Слово	F_1	F_2
1	2	3	4	1	2	3	4
Текст: XUD-01				Текст: XUD-17			
1	жер	7	3	1	көз	60	25
2	банды	6	3	2	қала	13	5
3	ел	5	3	3	мектеп	13	7
4	мылтық	4	1	4	жыл	8	7
5	әскер	3	1	5	жақ	7	6
6	ат	2	2	6	қыз	7	5
7	ауыл	2	1	7	үміт	7	4
8	жұрт	2	2	8	адам	6	5
9	жыл	2	2	9	дүние	6	5
10	коммунист	2	1	10	аспан	5	3
11	күз	2	2	11	бет-ауыз	5	5
12	кісі	2	1	12	ел	5	4
13	қазақ	2	1	13	күл	5	2
14	қатын-бала	2	1	14	қалып	5	3
15	қол	2	2	15	қол	5	4
16	орыс	2	1	16	от	5	2
17	поселке	2	1	-	- - - -	-	-
Текст: XUD-04				Текст: XUD-19			
1	Хамит	105	45	1	Жақсыбай	9	4
2	банды	66	21	2	Күнтілім	5	5
3	мылтық	50	18	3	үй	5	3
4	қол	45	18	4	бақыт	4	4
- -	- - - -	- -	- -	- -	- - - -	- -	- -
Текст: XUD-14				Текст: XUD-20			
1	Әміржан	22	16	1	Күнтілім	25	17
2	Әсия	21	14	2	бақыт	23	14
3	қыз	7	7	3	Ленин	15	10
4	бала	6	6	4	өмір	10	5
-	- - - -	-	-	-	- - - -	-	-

[4] F_1 – абсолютная частота слов в отдельном минитексте, F_2 – абсолютная частота абзацев в 20 минитекстах (число абзацев с данным словом).

Фрагмент 5

Потенциальный алфавитно-частотный словарь слов по научно-популярному типу текстов[5]

№ п/п	Слово	Абсолют. частота F	Число абзацев m	$K_{важ}$	ГОС, ВОС, ПС,
1	2	3	4	5	6
NAU-01; $K^1_{важ}$=0,00957; $K^2_{важ}$= 0,00426					
1	айырмашылық/зт	4	2	0,008510	ВОС
2	анықтал/ет	2	2	0,004255	ВОС
3	асбест/зт	3	1	0,003192	ПС
4	әдіс/зт	3	1	0,003192	ПС
5	байланысты/үс	2	2	0,004255	ВОС
6	байыт/ет	3	2	0,006382	ВОС
7	бол/ет	6	4	0,025532	ГОС
8	зат/зт	5	2	0,010638	ГОС
9	кейін/үс	2	2	0,004255	ВОС
10	кен/зт	5	3	0,015957	ГОС
11	кристалл/зт	10	3	0,031915	ГОС
12	күй/зт	2	2	0,004255	ВОС
13	қасиет/зт	6	3	0,019148	ГОС
14	минерал/зт	11	4	0,046808	ГОС
15	пайдаланыл/ет	2	2	0,004255	ВОС
16	түр/зт	2	2	0,004255	ВОС
NAU-02; $K^1_{важ}$=0,00045; $K^2_{важ}$=0,00019					
1	ажырат/ет	6	2	0,001000	ГОС
2	айнал/ет	2	2	0,000333	ВОС
3	айыр/ет	2	2	0,000333	ВОС
4	ал/ет	6	4	0,002000	ГОС
5	алын/ет	5	4	0,001667	ГОС
6	аппарат/зт	5	3	0,001250	ГОС
7	ара/зт	3	2	0,000500	ГОС
8	аралас/ет	2	2	0,000333	ВОС
9	арнаулы/сн	2	2	0,000333	ВОС
10	ас/ет	3	3	0,000750	ГОС
11	атал/ет	2	2	0,000333	ВОС
12	әсер/зт	4	2	0,000667	ГОС
13	байланысты/үс	3	3	0,000750	ГОС
14	байыт/ет	3	3	0,000750	ГОС
15	бар/ет	2	2	0,000333	ВОС
NAU-02; $K^1_{важ}$=0,00045; $K^2_{важ}$=0,00019					
16	бас/зт	4	1	0,000333	ВОС
17	бол/ет	7	4	0,002333	ГОС
18	болат/зт	2	2	0,000333	ВОС
19	бөл/ет	6	4	0,002000	ГОС
20	бөлік/зт	5	2	0,000833	ГОС
21	бөлін/ет	6	5	0,002500	ГОС
22	гидроциклон/зт	3	1	0,000250	ВОС
23	грохот/зт	3	1	0,000250	ВОС

[5] Для каждого слова словаря по заранее известным значениям $K^1_{важ}$, $K^2_{,важ}$ вычислено – $K_{важ}$ и установлена его принадлежность к ГОС, ВОС или ПС.

Продолжение фрагм. 5

1	2	3	4	5	6
24	даяр/сн	2	2	0,000333	ВОС
25	де/ет	5	5	0,002083	ГОС
26	диаметр/зт	3	3	0,000750	ГОС
27	диірмен/зт	13	4	0,004333	ГОС
28	ең/үс	4	4	0,001333	ГОС
29	ет/ет	5	4	0,001667	ГОС
30	жақ/зт	4	3	0,001000	ГОС
31	жат/ет	5	3	0,001250	ГОС
32	жер/зт	2	2	0,000333	ВОС
33	жет/ет	5	4	0,001250	ГОС
34	жоғары/үс	3	2	0,000500	ГОС
35	жұмыс/зт	4	3	0,001000	ГОС
36	жыныс/зт	3	3	0,000750	ГОС
37	кездес/ет	5	3	0,001250	ГОС
38	кейде/үс	2	2	0,000333	ВОС
39	кейін/үс	3	3	0,000750	ГОС
40	кел/ет	2	2	0,000333	ВОС
41	кен/зт	7	3	0,001750	ГОС
42	кесек/зт	9	6	0,004500	ГОС
43	классификатор/зт	4	2	0,000667	ГОС
44	көп/сн	5	4	0,001667	ГОС
45	күй/зт	2	2	0,000333	ВОС
46	күш/зт	4	3	0,001000	ГОС
47	қал/ет	3	3	0,000750	ГОС
48	қозғал/ет	2	2	0,000333	ВОС
49	қолданыл/ет	5	4	0,001667	ГОС
50	қуатты/сн	2	2	0,000333	ВОС
51	металл/зт	2	2	0,000333	ВОС
52	метр/зт	6	3	0,001500	ГОС
53	миллиметр/зт	5	4	0,001667	ГОС
54	минерал/зт	19	6	0,009500	ГОС
55	мөлшер/зт	5	5	0,002083	ГОС
56	мүмкіндік/зт	2	2	0,000333	ВОС
57	науа/зт	9	2	0,001500	ГОС
58	ондаған/үс	3	3	0,000750	ГОС
59	орын/зт	3	2	0,000500	ГОС
NAU-02; $K^1_{важ}$=0,00045; $K^2_{важ}$=0,00019					
60	осындай/сн	2	2	0,000333	ВОС
61	өткізіл/ет	3	2	0,000500	ГОС
62	пайдалы/сн	6	4	0,002000	ГОС
63	процесс/зт	7	6	0,003500	ГОС
64	пішін/зт	2	2	0,000333	ВОС
65	руда/зт	33	14	0,038500	ГОС
66	салмақ/зт	2	2	0,000333	ВОС
67	су/зт	5	3	0,001250	ГОС
68	тау/зт	3	3	0,000750	ГОС
69	темір/зт	2	2	0,000333	ВОС
70	тонна/зт	2	2	0,000333	ВОС
71	төмен/үс	2	2	0,000333	ВОС
72	тұр/ет	4	2	0,000667	ГОС
73	түйіршік/зт	8	5	0,003333	ГОС
74	түйіршікті/сн	7	3	0,001750	ГОС

Продолжение фрагм. 5

1	2	3	4	5	6
75	түп/зт	2	2	0,000333	ВОС
76	түр/зт	3	3	0,000750	ГОС
77	түрлі/сн	3	3	0,000750	ГОС
78	түс/ет	6	3	0,001500	ГОС
79	түсір/ет	2	2	0,000333	ВОС
80	түсіріл/ет	2	2	0,000333	ВОС
81	ұнтақта/ет	9	7	0,005250	ГОС
82	ұнтақтал/ет	3	1	0,000250	ВОС
83	ұсақ/сн	3	2	0,000500	ГОС
84	ұсақта/ет	2	2	0,000333	ВОС
85	ұсақтал/ет	3	2	0,000500	ГОС
86	ұсат/ет	5	4	0,001667	ГОС
87	ұсатқыш/зт	14	4	0,004667	ГОС
88	ұсатыл/ет	4	3	0,001000	ГОС
89	фабрика/зт	2	2	0,000333	ВОС
90	шар/зт	4	1	0,000333	ВОС
91	шық/ет	4	2	0,000667	ГОС
92	ірі/сн	6	4	0,002000	ГОС
93	ірілік/зт	2	2	0,000333	ВОС
94	іш/зт	5	4	0,001667	ГОС
95	электр/зт	2	2	0,000333	ВОС
96	энергия/зт	6	5	0,002500	ГОС
NAU-03; $K^1_{важ}$= 0,03061; $K^2_{важ}$= 0,02041					
1	әдіс/зт	8	2	0,16326	ГОС
2	байыту/ет	5	2	0,10204	ГОС
3	гравитациялық/сн	5	2	0,10204	ГОС
4	кен/зт	2	1	0,02041	ВОС
5	қолданыл/ет	3	2	0,06122	ГОС
NAU-04; $K^1_{важ}$=0,00248; $K^2_{важ}$=0,00115					
1	ағыз/ет	4	2	0,00263	ГОС
2	аз/сн	2	2	0,00131	ВОС
3	ақ/ет	2	2	0,00131	ВОС
4	алға/үс	2	2	0,00131	ВОС
5	арнаулы/сн	2	2	0,00131	ВОС
6	әрекет/зт	3	1	0,00098	ПС
7	әсер/зт	7	2	0,00456	ГОС
8	бағыт/зт	14	2	0,00919	ГОС
9	бас/зт	2	2	0,00131	ВОС
10	бет/зт	8	4	0,01051	ГОС
11	бой/зт	4	2	0,00263	ГОС
12	бол/ет	7	4	0,00919	ГОС
13	бөлін/ет	3	2	0,00197	ВОС
14	де/ет	3	3	0,00295	ГОС
15	дека/зт	8	2	0,00525	ГОС
16	екпінді/сн	6	2	0,00394	ГОС
17	ет/ет	3	2	0,00197	ВОС
18	жет/ет	3	3	0,00295	ГОС
19	жоғары/үс	2	2	0,00131	ВОС
20	жол/зт	2	2	0,00131	ВОС
21	жылжы/ет	4	2	0,00263	ГОС
22	көлденең/сн	3	2	0,00197	ВОС
23	күш/зт	21	2	0,01379	ГОС

Окончание фрагм. 5

1	2	3	4	5	6
24	қабаттағы/сн	2	2	0,00131	ВОС
25	қозғал/ет	4	1	0,00131	ВОС
26	қозғалыс/зт	5	1	0,00164	ВОС
27	машина/зт	3	1	0,00098	ПС
28	метр/зт	2	1	0,00066	ПС
29	миллиметр/зт	2	1	0,00066	ПС
30	минерал/зт	16	6	0,03153	ГОС
31	науашық/зт	2	2	0,00131	ВОС
32	орналас/ет	2	2	0,00131	ВОС
33	осылайша/үс	2	2	0,00131	ВОС
34	осындай/сн	3	2	0,00197	ВОС
35	процесс/зт	5	3	0,00493	ГОС
36	салмақ/зт	3	3	0,00295	ГОС
37	стол/зт	13	3	0,01281	ГОС
38	су/зт	7	2	0,00460	ГОС
39	тербеліс/зт	2	2	0,00131	ВОС
40	түйіршік/зт	15	4	0,01970	ГОС
41	түрлі/сн	7	3	0,00690	ГОС
42	түс/ет	4	3	0,00390	ГОС
43	ұзына/үс	4	2	0,00263	ГОС
44	үлесті/сн	3	3	0,00295	ГОС

Список 1. Аффиксы образования словоформ от корневых и производных основ существительных с конечным мягким слогом[6]

Возможные типы слов-основ и конечных звуков (букв) у существительных в реестре **АСКЯ**					
№ п/п	Тип 1: на гласные (кроме ***у, и***)	Тип 2: на ***у, и, й, р***	Тип 3: на ***м, н, ң***	Тип 4: на ***б, в, г, д, п, т, к, ф, ш, с, қ, х, ч, ц, һ***	Тип 5: на ***ж, з, л***
	Әже, келі, мәсі, мекеме	*Үй, белбеу, темір, сәби*	*Кілем, кескін, өлең*	*Шөп, сүт, күрек, жүріс*	*Тәж, жез, киіз, көз, ел*
1	2	3	4	5	6
1	*ге*	*ге*	*ге*	*ке*	*ге*
2	*де*	*де*	*де*	*те*	*де*
3	*ден*	*ден*	*нен*	*тен*	*ден*
4	*лер*	*лер*	*дер*	*тер*	*дер*
5	*лерге*	*лерге*	*дерге*	*терге*	*дерге*
6	*лерде*	*лерде*	*дерде*	*терде*	*дерде*
7	*лерден*	*лерден*	*дерден*	*терден*	*дерден*
8	*лерді*	*лерді*	*дерді*	*терді*	*дерді*
9	*лердің*	*лердің*	*дердің*	*тердің*	*дердің*
10	*лермен*	*лермен*	*дермен*	*термен*	*дермен*
11	*лерменен*	*лерменен*	*дерменен*	*терменен*	*дерменен*
12	*лері*	*лері*	*дері*	*тері*	*дері*
13	*лерім*	*лерім*	*дерім*	*терім*	*дерім*
14	*лерімде*	*лерімде*	*дерімде*	*терімде*	*дерімде*
15	*лерімді*	*лерімді*	*дерімді*	*терімді*	*дерімді*
16	*леріме*	*леріме*	*деріме*	*теріме*	*деріме*
17	*лерімен*	*лерімен*	*дерімен*	*терімен*	*дерімен*
18	*леріменен*	*леріменен*	*деріменен*	*теріменен*	*деріменен*
19	*леріммен*	*леріммен*	*дерімме*н	*теріммен*	*дерімме*н
20	*лерімменен*	*лерімменен*	*дерімменен*	*терімменен*	*дерімменен*
21	*лерімнен*	*лерімнен*	*длерімнен*	*терімнен*	*длерімнен*
22	*лерімнің*	*лерімнің*	*дерімнің*	*терімнің*	*дерімнің*
23	*леріміз*	*леріміз*	*деріміз*	*теріміз*	*деріміз*
24	*лерімізбен*	*лерімізбен*	*дерімізбен*	*терімізбен*	*дерімізбен*

[6] В списке 1 аффиксы сказуемости могут отсутствовать. Аффиксы в колонках в большинстве случаев могут соответствовать друг-другу, а иногда отличаться. Такой подход к составлению списка аффиксов позволяеет избежать многократные проверки в процессе синтеза именных словоформ в алгоритме программы.

1	2	3	4	5	6
25	*лерімізбенен*	*лерімізбенен*	*дерімізбенен*	*терімізбенен*	*дерімізбенен*
26	*лерімізге*	*лерімізге*	*дерімізге*	*терімізге*	*дерімізге*
27	*лерімізде*	*лерімізде*	*дерімізде*	*терімізде*	*дерімізде*
28	*лерімізден*	*лерімізден*	*дерімізден*	*терімізден*	*дерімізден*
29	*лерімізді*	*лерімізді*	*дерімізді*	*терімізді*	*дерімізді*
30	*леріміздің*	*леріміздің*	*деріміздің*	*теріміздің*	*деріміздің*
31	*лерінде*	*лерінде*	*дерінде*	*терінде*	*дерінде*
32	*леріне*	*леріне*	*деріне*	*теріне*	*деріне*
33	*лерінен*	*лерінен*	*дерінен*	*терінен*	*дерінен*
34	*лерінің*	*лерінің*	*дерінің*	*терінің*	*дерінің*
35	*лерің*	*лерің*	*дерің*	*терің*	*дерің*
36	*леріңде*	*леріңде*	*деріңде*	*теріңде*	*деріңде*
37	*леріңді*	*леріңді*	*деріңді*	*теріңді*	*деріңді*
38	*леріңе*	*леріңе*	*деріңе*	*теріңе*	*деріңе*
39	*леріңмен*	*леріңмен*	*деріңмен*	*теріңмен*	*деріңмен*
40	*леріңменен*	*леріңменен*	*деріңменен*	*теріңменен*	*деріңменен*
41	*леріңнен*	*леріңнен*	*деріңнен*	*теріңнен*	*деріңнен*
42	*леріңнің*	*леріңнің*	*деріңнің*	*теріңнің*	*деріңнің*
43	*леріңіз*	*леріңіз*	*деріңіз*	*теріңіз*	*деріңіз*
44	*леріңізбен*	*леріңізбен*	*деріңізбен*	*теріңізбен*	*деріңізбен*
45	*леріңізбенен*	*леріңізбенен*	*деріңізбенен*	*теріңізбенен*	*деріңізбенен*
46	*леріңізге*	*леріңізге*	*деріңізге*	*теріңізге*	*деріңізге*
47	*леріңізде*	*леріңізде*	*деріңізде*	*теріңізде*	*деріңізде*
48	*леріңізден*	*леріңізден*	*деріңізден*	*теріңізден*	*деріңізден*
49	*леріңізді*	*леріңізді*	*деріңізді*	*теріңізді*	*деріңізді*
50	*леріңіздің*	*леріңіздің*	*деріңіздің*	*теріңіздің*	*еріңіздің*
51	*м*	*ім*	*ім*	*ім*	*ім*
52	*мде*	*імде*	*імде*	*імде*	*імде*
53	*мді*	*імді*	*імді*	*імді*	*імді*
54	*мдікі*	*імдікі*	*імдікі*	*імдікі*	*імдікі*
55	*мдікімен*	*імдікімен*	*імдікімен*	*імдікімен*	*імдікімен*
56	*мдікіменен*	*імдікіменен*	*імдікіменен*	*імдікіменен*	*імдікіменен*

Продолжение спис. 1

1	2	3	4	5	6
57	мдікіне	імдікіне	імдікіне	імдікіне	імдікіне
58	мдікінен	імдікінен	імдікінен	імдікінен	імдікінен
59	мдікінің	імдікінің	імдікінің	імдікінің	імдікінің
60	ме	іме	іме	ме	іме
61	мен	мен	мен	пен	бен ("мен" для "л")
62	менен	менен	менен	пенен	бенен ("менен"для"л")
63	ммен	іммен	іммен	іммен	іммен
64	мменен	імменен	імменен	імменен	імменен
65	мнен	імнен	імнен	імнен	імнен
66	мнің	імнің	імнің	мнің	імнің
67	міз	іміз	іміз	іміз	іміз
68	мізбен	імізбен	імізбен	імізбен	імізбен
69	мізбенен	імізбенен	імізбенен	імізбенен	імізбенен
70	мізге	імізге	імізге	імізге	імізге
71	мізде	імізде	імізде	імізде	імізде
72	мізден	імізден	імізден	імізден	імізден
73	мізді	імізді	імізді	імізді	імізді
74	міздікі	іміздікі	іміздікі	іміздікі	іміздікі
75	міздікімен	іміздікімен	іміздікімен	іміздікімен	іміздікімен
76	міздікіменен	іміздікіменен	іміздікіменен	іміздікіменен	іміздікіменен
77	міздікіне	іміздікіне	іміздікіне	іміздікіне	іміздікіне
78	міздікінен	іміздікінен	іміздікінен	іімміздікінен	іміздікінен
79	міздікінің	іміздікінің	іміздікінің	іміздікінің	іміздікінің
80	міздің	іміздің	іміздің	іміздің	іміздің
81	ні	ді	ді	ті	ді
82	нікі	дікі	дікі	тікі	дікі
83	нікіне	дікіне	дікіне	тікіне	дікіне
84	нікімен	дікімен	дікімен	тікімен	дікімен
85	нікіменен	дікіменен	дікіменен	тікіменен	дікіменен
86	нікінен	дікінен	дікінен	тікінен	дікінен
87	нікінің	дікінің	дікінің	тікінің	дікінің
88	нің	дің	нің	тің	дің
89	ң	ің	ің	ің	ің

1	2	3	4	5	6
90	*ңде*	*іңде*	*іңде*	*іңде*	*іңде*
91	*ңді*	*іңді*	*іңді*	*іңді*	*іңді*
92	*ңдікі*	*іңдікі*	*іңдікі*	*іңдікі*	*іңдікі*
93	*ңдікімен*	*іңдікімен*	*іңдікімен*	*іңдікімен*	*іңдікімен*
94	*ңдікіменен*	*іңдікіменен*	*іңдікіменен*	*іңдікіменен*	*іңдікіменен*
95	*ңдікіне*	*іңдікіне*	*іңдікіне*	*іңдікіне*	*іңдікіне*
96	*ңдікінен*	*іңдікінен*	*іңдікінен*	*іңдікінен*	*іңдікінен*
97	*ңдікінің*	*іңдікінің*	*іңдікінің*	*іңдікінің*	*іңдікінің*
98	*ңе*	*іңе*	*іңе*	*іңе*	*іңе*
99	*ңмен*	*іңмен*	*іңмен*	*іңмен*	*іңмен*
100	*ңменен*	*іңменен*	*іңменен*	*іңменен*	*іңменен*
101	*ңмін*	*іңмін*	*іңмін*	*іңмін*	*іңмін*
102	*ңнен*	*іңнен*	*іңнен*	*іңнен*	*іңнен*
103	*ңнің*	*іңнің*	*іңнің*	*іңнің*	*іңнің*
104	*ңіз*	*іңіз*	*іңіз*	*іңіз*	*іңіз*
105	*ңізбен*	*іңізбен*	*іңізбен*	*іңізбен*	*іңізбен*
106	*ңізбенен*	*іңізбенен*	*іңізбенен*	*іңізбенен*	*іңізбенен*
107	*ңізге*	*іңізге*	*іңізге*	*іңізге*	*іңізге*
108	*ңізде*	*іңізде*	*іңізде*	*іңізде*	*іңізде*
109	*ңізден*	*іңізден*	*іңізден*	*іңізден*	*іңізден*
110	*ңізді*	*іңізді*	*іңізді*	*іңізді*	*іңізді*
111	*ңіздікі*	*іңіздікі*	*іңіздікі*	*іңіздікі*	*іңіздікі*
112	*ңіздікімен*	*іңіздікімен*	*іңіздікімен*	*іңіздікімен*	*іңіздікімен*
113	*ңіздікіменен*	*іңіздікіменен*	*іңіздікіменен*	*іңіздікіменен*	*іңіздікіменен*
114	*ңіздікіне*	*іңіздікіне*	*іңіздікіне*	*іңіздікіне*	*іңіздікіне*
115	*ңіздікінен*	*іңіздікінен*	*іңіздікінен*	*іңіздікінен*	*іңіздікінен*
116	*ңіздікінің*	*іңіздікінің*	*іңіздікінің*	*іңіздікінің*	*іңіздікінің*
117	*ңіздің*	*іңіздің*	*іңіздің*	*іңіздің*	*іңіздің*
118	*сі*	*і*	*і*	*і*	*і*
119	*сімен*	*імен*	*імен*	*імен*	*імен*
120	*сіменен*	*іменен*	*іменен*	*іменен*	*іменен*
121	*сінде*	*інде*	*інде*	*інде*	*інде*
122	*сіне*	*іне*	*іне*	*іне*	*іне*

Окончание спис. 1

1	2	3	4	5	6
123	*сінен*	*інен*	*інен*	*інен*	*інен*
124	*сінікі*	*інікі*	*інікі*	*інікі*	*інікі*
125	*сінікімен*	*інікімен*	*інікімен*	*інікімен*	*інікімен*
126	*сінікіменен*	*інікіменен*	*інікіменен*	*інікіменен*	*інікіменен*
127	*сінікіне*	*інікіне*	*інікіне*	*інікіне*	*інікіне*
128	*сінікінен*	*інікінен*	*інікінен*	*інікінен*	*інікінен*
129	*сінікінің*	*інікінің*	*інікінің*	*інікінің*	*інікінің*
130	*сінің*	*інің*	*інің*	*інің*	*інің*

Список 2. **Аффиксы образования словоформ от корневых и производных основ существительных с конечным мягким слогом[7]**

Возможные типы слов-основ и конечных звуков (букв) у существительных в реестре **АСКЯ**				
№ *п/п*	Тип 1: на гласные, кроме ***у, и***	Тип 2: на ***у, и, й, р***	Тип 3: на ***м, н, ң, ж, з, л***	Тип 4: на: ***в, г, д, п, т, к, ф, ш, с, қ, х, ч, ц, һ***
	кісі, кенже, кәсіпші, ене, биші	*дәрігер, сәби, керей, кедей, күйеу*	*көсем, тәж, келін, кербез, әйел*	*жігіт*
1	2	3	4	5
1	*лерміз*	*лерміз*	*дерміз*	*терміз*
2	*лермісіздер*	*лермісіздер*	*дермісіздер*	*термісіздер*
3	*лерсіз*	*лерсіз*	*дерсіз*	*терсіз*
4	*лерсіздер*	*лерсіздер*	*дерсіздер*	*терсіздер*
5	*лерсіңдер*	*лерсіңдер*	*дерсіңдер*	*терсіңдер*
6	*лерімбісіздер*	*лерімбісіздер*	*дерімбісіздер*	*терімбісіздер*
7	*лерімбісіңдер*	*лерімбісіңдер*	*дерімбісіңдер*	*терімбісіңдер*
8	*лерімсіздер*	*лерімсіздер*	*дерімсіздер*	*терімсіздер*
9	*лерімсіңдер*	*лерімсіңдер*	*дерімсіңдер*	*терімсіңдер*
10	*леріміз*	*леріміз*	*деріміз*	*теріміз*
11	*лерімісіз*	*лерімісіз*	*дерімісіз*	*терімісіз*
12	*лерімісіздер*	*лерімісіздер*	*дерімісіздер*	*терімісіздер*
13	*лерімісіңдер*	*лерімісіңдер*	*дерімісіңдер*	*терімісіңдер*
14	*леріңізбіз*	*леріңізбіз*	*деріңізбіз*	*теріңізбіз*
15	*лерісіз*	*лерісіз*	*дерісіз*	*терісіз*
16	*лерісіздер*	*лерісіздер*	*дерісіздер*	*терісіздер*
17	*лерісіңдер*	*лерісіңдер*	*дерісіңдер*	*терісіңдер*
18	*мбісің*	*імбісің*	*імбісің*	*імбісің*
19	*мсіз*	*імсіз*	*імсіз*	*імсіз*
20	*мсіздер*	*імсіздер*	*імсіздер*	*імсіздер*
21	*мсің*	*імсің*	*імсің*	*імсің*
22	*мсіңдер*	*імсіңдер*	*імсіңдер*	*імсіңдер*
23	*мізбісіз*	*імізбісіз*	*імізбісіз*	*імізбісіз*

[7] В списке обязательно должны участвовать аффиксы сказуемости. Аффиксы в колонках в большинстве случаев могут соответствовать друг-другу, а иногда отличаться. Такой подход к составлению списка аффиксов позволяеет избежать многократные проверки в процессе синтеза именных словоформ в алгоритме программы.

Окончание спис. 2

1	2	3	4	5
24	*мізсіз*	*імізсіз*	*імізсіз*	*імізсіз*
25	*мізсіздер*	*імізсіздер*	*імізсіздер*	*імізсіздер*
26	*мізсіңдер*	*імізсіңдер*	*імізсіңдер*	*імізсіңдер*
27	*мін*	*мін*	*мін* (если ж, з, то *бін*)	*пін*
28	*місіз*	*місіз*	*бісіз* (“л” - *місіз*)	*пісіз*
29	*місіздер*	*місіздер*	*бісіздер* (“л” - *місіздер*)	*пісіздер*
30	*місің*	*місің*	*бісің* (“л” - *місің*)	*пісің*
31	*місіңдер*	*місіңдер*	*бісіңдер* (“л” - *місіңдер*)	*пісіңдер*
32	*ңбіз*	*іңбіз*	*іңбіз*	*іңбіз*
33	*ңмін*	*іңмін*	*іңмін*	*іңмін*
34	*ңізбіз*	*іңізбіз*	*іңізбіз*	*іңізбіз*
35	*ңізбін*	*іңізбін*	*іңізбін*	*іңізбін*
36	*сіз*	*сіз*	*сіз*	*сіз*
37	*сіздер*	*сіздер*	*сіздер*	*сіздер*
38	*сімін*	*імін*	*імін*	*імін*
39	*сімісіз*	*імісіз*	*імісіз*	*імісіз*
40	*сімісіздер*	*імісіздер*	*імісіздер*	*імісіздер*
41	*сімісің*	*імісің*	*імісің*	*імісің*
42	*сің*	*сің*	*сің*	*сің*
43	*сіңдер*	*сіңдер*	*сіңдер*	*сіңдер*
44	*сісіз*	*ісіз*	*ісіз*	*ісіз*
45	*сісіздер*	*ісіздер*	*ісіздер*	*ісіздер*
46	*сісің*	*ісің*	*ісің*	*ісің*
47	*сісіңдер*	*ісіңдер*	*ісіңдер*	*ісіңдер*

Список 3. **Аффиксы синтеза словоформ от корневых и производных глагольных основ с конечным твердым слогом[8]**

Возможные типы слов и конечных слогов в глагольной основе (в реестре автоматического словаря)				
№ п/п	Тип 1: *Айт+атын, айт+пақ, айт+қанша...*	Тип 2: *Бар+атын, бар+мақ, бар+ғанша ...*	Тип 3: *Қон+атын, қон+бақ, он+ғанша...*	Тип 4: *Тала+йтын, тала+мақ, тала+ғанша*
1	2	3	4	5
1	*а*	*а*	*а*	-
2	*ады*	*ады*	*ады*	*ды*
3	*айық*	*айық*	*айық*	*йық*
4	*айықшы*	*айықшы*	*айықшы*	*йықшы*
5	*айын*	*айын*	*айын*	*йын*
6	*айыншы*	*айыншы*	*айыншы*	*йыншы*
7	*ам*	*ам*	*ам*	-
8	*амыз*	*амыз*	*амыз*	*ймыз*
9	*амын*	*амын*	*амын*	*ймын*
10	*ар*	*ар*	*ар*	*р*
11	*арға*	*арға*	*арға*	*рға*
12	*арда*	*арда*	*арда*	*рда*
13	*ардай*	*ардай*	*ардай*	*рдай*
14	*ардан*	*ардан*	*ардан*	*рдан*
15	*арды*	*арды*	*арды*	*рды*
16	*ардың*	*ардың*	*ардың*	*рдың*
17	*армен*	*армен*	*армен*	*рмен*
18	*арменен*	*арменен*	*арменен*	*рменен*
19	*армыз*	*армыз*	*армыз*	*рмыз*
20	*армын*	*армын*	*армын*	*рмын*
21	*арсыз*	*арсыз*	*арсыз*	*рсыз*
22	*арсыздар*	*арсыздар*	*арсыздар*	*рсыздар*
23	*арсың*	*арсың*	*арсың*	*рсың*
24	*арсыңдар*	*арсыңдар*	*арсыңдар*	*рсыңдар*
25	*ары*	*ары*	*ары*	*ры*

[8] Аффиксы в колонках в большинстве случаев могут соответствовать друг-другу, а иногда отличаться. Такой подход к составлению списка аффиксов позволяеет избежать многократные проверки в процессе синтеза глагольных словоформ в алгоритме программы.

Продолжение спис. 3

1	2	3	4	5
26	арым	арым	арым	рым
27	арыма	арыма	арыма	рыма
28	арымда	арымда	арымда	рымда
29	арымдай	арымдай	арымдай	рымдай
30	арымды	арымды	арымды	рымды
31	арымен	арымен	арымен	рымен
32	арыменен	арыменен	арыменен	рыменен
33	арыммен	арыммен	арыммен	рыммен
34	арымменен	арымменен	арымменен	рымменен
35	арымнан	арымнан	арымнан	рымнан
36	арымның	арымның	арымның	рымның
37	арымыз	арымыз	арымыз	рымыз
38	арымызбен	арымызбен	арымызбен	рымызбен
39	арымызбенен	арымызбенен	арымызбенен	рымызбенен
40	арымызга	арымызга	арымызга	рымызга
41	арымызда	арымызда	арымызда	рымызда
42	арымыздай	арымыздай	арымыздай	рымыздай
43	арымыздан	арымыздан	арымыздан	рымыздан
44	арымызды	арымызды	арымызды	рымызды
45	арымыздың	арымыздың	арымыздың	рымыздың
46	арына	арына	арына	рына
47	арынан	арынан	арынан	рынан
48	арында	арында	арында	рында
49	арының	арының	арының	рының
50	арың	арың	арың	рың
51	арыңа	арыңа	арыңа	рыңа
52	арыңда	арыңда	арыңда	рыңда
53	арыңдай	арыңдай	арыңдай	рыңдай
54	арыңды	арыңды	арыңды	рыңды
55	арыңмен	арыңмен	арыңмен	рыңмен
56	арыңменен	арыңменен	арыңменен	рыңменен
57	арыңнан	арыңнан	арыңнан	рыңнан
58	арыңның	арыңның	арыңның	рыңның

Продолжение спис. 3

1	2	3	4	5
59	*арыңыз*	*арыңыз*	*арыңыз*	*рыңыз*
60	*арыңызбен*	*арыңызбен*	*арыңызбен*	*рыңызбен*
61	*арыңызбенен*	*арыңызбенен*	*арыңызбенен*	*рыңызбенен*
62	*арыңызга*	*арыңызга*	*арыңызга*	*рыңызга*
63	*арыңызда*	*арыңызда*	*арыңызда*	*рыңызда*
64	*арыңыздай*	*арыңыздай*	*арыңыздай*	*рыңыздай*
65	*арыңыздан*	*арыңыздан*	*арыңыздан*	*рыңыздан*
66	*арыңызды*	*арыңызды*	*арыңызды*	*рыңызды*
67	*арыңыздың*	*арыңыздың*	*арыңыздың*	*рыңыздың*
68	*асыз*	*асыз*	*асыз*	*йсыз*
69	*асыздар*	*асыздар*	*асыздар*	*йсыздар*
70	*асың*	*асың*	*асың*	*йсың*
71	*асыңдар*	*асыңдар*	*асыңдар*	*йсыңдар*
72	*ататындарындай*	*ататындарындай*	*ататындарындай*	*татындарындай*
73	*атын*	*атын*	*атын*	*йтын*
74	*атынбыз*	*атынбыз*	*атынбыз*	*йтынбыз*
75	*атынга*	*атынга*	*атынга*	*йтынга*
76	*атында*	*атында*	*атында*	*йтында*
77	*атындай*	*атындай*	*атындай*	*йтындай*
78	*атындар*	*атындар*	*атындар*	*йтындар*
79	*атындарга*	*атындарга*	*атындарга*	*йтындарга*
80	*атындарда*	*атындарда*	*атындарда*	*йтындарда*
81	*атындардай*	*атындардай*	*атындардай*	*йтындардай*
82	*атындардан*	*атындардан*	*атындардан*	*йтындардан*
83	*атындарды*	*атындарды*	*атындарды*	*йтындарды*
84	*атындардың*	*атындардың*	*атындардың*	*йтындардың*
85	*атындармен*	*атындармен*	*атындармен*	*йтындармен*
86	*атындарменен*	*атындарменен*	*атындарменен*	*йтындарменен*
87	*атындарымен*	*атындарымен*	*атындарымен*	*йтындарымен*
88	*атындарыменен*	*атындарыменен*	*атындарыменен*	*йтындарыменен*
89	*атындарыммен*	*атындарыммен*	*атындарыммен*	*йтындарыммен*
90	*атындарымменен*	*атындарымменен*	*атындарымменен*	*йтындарымменен*
91	*атындарымнан*	*атындарымнан*	*атындарымнан*	*йтындарымнан*

Продолжение спис. 3

1	2	3	4	5
92	атындарымның	атындарымның	атындарымның	йтындарымның
93	атындарымыз	атындарымыз	атындарымыз	йтындарымыз
94	атындарымызбен	атындарымызбен	атындарымызбен	йтындарымызбен
95	атындарымызбенен	атындарымызбенен	атындарымызбенен	йтындарымызбенен
96	атындарымызга	атындарымызга	атындарымызга	йтындарымызга
97	атындарымызда	атындарымызда	атындарымызда	йтындарымызда
98	атындарымыздай	атындарымыздай	атындарымыздай	йтындарымыздай
99	атындарымыздан	атындарымыздан	атындарымыздан	йтындарымыздан
100	атындарымызды	атындарымызды	атындарымызды	йтындарымызды
101	атындарымыздың	атындарымыздың	атындарымыздың	йтындарымыздың
102	атындарын	атындарын	атындарын	йтындарын
103	атындарына	атындарына	атындарына	йтындарына
104	атындарынан	атындарынан	атындарынан	йтындарынан
105	атындарында	атындарында	атындарында	йтындарында
106	атындарының	атындарының	атындарының	йтындарының
107	атындарың	атындарың	атындарың	йтындарың
108	атындарыңа	атындарыңа	атындарыңа	йтындарыңа
109	атындарыңда	атындарыңда	атындарыңда	йтындарыңда
110	атындарыңдай	атындарыңдай	атындарыңдай	йтындарыңдай
111	атындарыңды	атындарыңды	атындарыңды	йтындарыңды
112	атындарыңмен	атындарыңмен	атындарыңмен	йтындарыңмен
113	атындарыңменен	атындарыңменен	атындарыңменен	йтындарыңменен
114	атындарыңнан	атындарыңнан	атындарыңнан	йтындарыңнан
115	атындарыңның	атындарыңның	атындарыңның	йтындарыңның
116	атындарыңыз	атындарыңыз	атындарыңыз	йтындарыңыз
117	атындарыңызбен	атындарыңызбен	атындарыңызбен	йтындарыңызбен
118	атындарыңызбенен	атындарыңызбенен	атындарыңызбенен	йтындарыңызбенен
119	атындарыңызга	атындарыңызга	атындарыңызга	йтындарыңызга
120	атындарыңызда	атындарыңызда	атындарыңызда	йтындарыңызда
121	атындарыңыздай	атындарыңыздай	атындарыңыздай	йтындарыңыздай
122	атындарыңыздан	атындарыңыздан	атындарыңыздан	йтындарыңыздан
123	атындарыңызды	атындарыңызды	атындарыңызды	йтындарыңызды
124	атындарыңыздың	атындарыңыздың	атындарыңыздың	йтындарыңыздың

1	2	3	4	5
125	*атынды*	*атынды*	*атынды*	*йтынды*
126	*атынмен*	*атынмен*	*атынмен*	*йтынмен*
127	*атынменен*	*атынменен*	*атынменен*	*йтынменен*
128	*атынмын*	*атынмын*	*атынмын*	*йтынмын*
129	*атыннан*	*атыннан*	*атыннан*	*йтыннан*
130	*атынның*	*атынның*	*атынның*	*йтынның*
131	*атынсыз*	*атынсыз*	*атынсыз*	*йтынсыз*
132	*атынсың*	*атынсың*	*атынсың*	*йтынсың*
133	*атыным*	*атыным*	*атыным*	*йтыным*
134	*атыныма*	*атыныма*	*атыныма*	*йтыныма*
135	*атынымда*	*атынымда*	*атынымда*	*йтынымда*
136	*атынымдай*	*атынымдай*	*атынымдай*	*йтынымдай*
137	*атынымды*	*атынымды*	*атынымды*	*йтынымды*
138	*атыныммен*	*атыныммен*	*атыныммен*	*йтыныммен*
139	*атынымменен*	*атынымменен*	*атынымменен*	*йтынымменен*
140	*атынымнан*	*атынымнан*	*атынымнан*	*йтынымнан*
141	*атынымның*	*атынымның*	*атынымның*	*йтынымның*
142	*атынымыз*	*атынымыз*	*атынымыз*	*йтынымыз*
143	*атынымызбен*	*атынымызбен*	*атынымызбен*	*йтынымызбен*
144	*атынымызбенен*	*атынымызбенен*	*атынымызбенен*	*йтынымызбенен*
145	*атынымызга*	*атынымызга*	*атынымызга*	*йтынымызга*
146	*атынымызда*	*атынымызда*	*атынымызда*	*йтынымызда*
147	*атынымыздай*	*атынымыздай*	*атынымыздай*	*йтынымыздай*
148	*атынымыздан*	*атынымыздан*	*атынымыздан*	*йтынымыздан*
149	*атынымызды*	*атынымызды*	*атынымызды*	*атынымызды*
150	*атынымыздың*	*атынымыздың*	*атынымыздың*	*атынымыздың*
151	*атының*	*атының*	*атының*	*йтының*
152	*атыныңа*	*атыныңа*	*атыныңа*	*йтыныңа*
153	*атыныңда*	*атыныңда*	*атыныңда*	*йтыныңда*
154	*атыныңдай*	*атыныңдай*	*атыныңдай*	*йтыныңдай*
155	*атыныңды*	*атыныңды*	*атыныңды*	*йтыныңды*
156	*атыныңмен*	*атыныңмен*	*атыныңмен*	*йтыныңмен*
157	*атыныңменен*	*атыныңменен*	*атыныңменен*	*йтыныңменен*

Продолжение спис. 3

1	2	3	4	5
158	атыныңнан	атыныңнан	атыныңнан	йтыныңнан
159	атыныңның	атыныңның	атыныңның	йтыныңның
160	атыныңыз	атыныңыз	атыныңыз	йтыныңыз
161	атыныңызбен	атыныңызбен	атыныңызбен	йтыныңызбен
162	атыныңызбенен	атыныңызбенен	атыныңызбенен	йтыныңызбенен
163	атыныңызға	атыныңызға	атыныңызға	йтыныңызға
164	атыныңызда	атыныңызда	атыныңызда	йтыныңызда
165	атыныңыздай	атыныңыздай	атыныңыздай	йтыныңыздай
166	атыныңыздан	атыныңыздан	атыныңыздан	йтыныңыздан
167	атыныңызды	атыныңызды	атыныңызды	йтыныңызды
168	атыныңыздың	атыныңыздың	атыныңыздың	йтыныңыздың
169	қай	ғай	ғай	ғай
170	қаймын	ғаймын	ғаймын	ғаймын
171	қайсыз	ғайсыз	ғайсыз	ғайсыз
172	қайсың	ғайсың	ғайсың	ғайсың
173	қалы	ғалы	ғалы	ғалы
174	қан	ған	ған	ған
175	қанбыз	ғанбыз	ғанбыз	ғанбыз
176	қанға	ғанға	ғанға	ғанға
177	қанда	ғанда	ғанда	ғанда
178	қандай	ғандай	ғандай	ғандай
179	қандайға	ғандайға	ғандайға	ғандайға
180	қандайда	ғандайда	ғандайда	ғандайда
181	қандайдан	ғандайдан	ғандайдан	ғандайдан
182	қандайды	ғандайды	ғандайды	ғандайды
183	қандайдың	ғандайдың	ғандайдың	ғандайдың
184	қандаймен	ғандаймен	ғандаймен	ғандаймен
185	қандайменен	ғандайменен	ғандайменен	ғандайменен
186	қандаймыз	ғандаймыз	ғандаймыз	ғандаймыз
187	қандаймын	ғандаймын	ғандаймын	ғандаймын
188	қандайсыз	ғандайсыз	ғандайсыз	ғандайсыз
189	қандайсың	ғандайсың	ғандайсың	ғандайсың
190	қандайым	ғандайым	ғандайым	ғандайым

1	2	3	4	5
191	қандайымдай	ғандайымдай	ғандайымдай	ғандайымдай
192	қандайымыз	ғандайымыз	ғандайымыз	ғандайымыз
193	қандайымыздай	ғандайымыздай	ғандайымыздай	ғандайымыздай
194	қандайың	ғандайың	ғандайың	ғандайың
195	қандайыңдай	ғандайыңдай	ғандайыңдай	ғандайыңдай
196	қандайыңыз	ғандайыңыз	ғандайыңыз	ғандайыңыз
197	қандайыңыздай	ғандайыңыздай	ғандайыңыздай	ғандайыңыздай
198	қандар	ғандар	ғандар	ғандар
199	қандарға	ғандарға	ғандарға	ғандарға
200	қандарда	ғандарда	ғандарда	ғандарда
201	қандардай	ғандардай	ғандардай	ғандардай
202	қандардан	ғандардан	ғандардан	ғандардан
203	қандарды	ғандарды	ғандарды	ғандарды
204	қандардың	ғандардың	ғандардың	ғандардың
205	қандармен	ғандармен	ғандармен	ғандармен
206	қандарменен	ғандарменен	ғандарменен	ғандарменен
207	қандары	ғандары	ғандары	ғандары
208	қандарым	ғандарым	ғандарым	ғандарым
209	қандарыма	ғандарыма	ғандарыма	ғандарыма
210	қандарымда	ғандарымда	ғандарымда	ғандарымда
211	қандарымдай	ғандарымдай	ғандарымдай	ғандарымдай
212	қандарымды	ғандарымды	ғандарымды	ғандарымды
213	қандарымен	ғандарымен	ғандарымен	ғандарымен
214	қандарыменен	ғандарыменен	ғандарыменен	ғандарыменен
215	қандарыммен	ғандарыммен	ғандарыммен	ғандарыммен
216	қандарымменен	ғандарымменен	ғандарымменен	ғандарымменен
217	қандарымнан	ғандарымнан	ғандарымнан	ғандарымнан
218	қандарымның	ғандарымның	ғандарымның	ғандарымның
219	қандарымыз	ғандарымыз	ғандарымыз	ғандарымыз
220	қандарымызбен	ғандарымызбен	ғандарымызбен	ғандарымызбен
221	қандарымызбенен	ғандарымызбенен	ғандарымызбенен	ғандарымызбенен
222	қандарымызға	ғандарымызға	ғандарымызға	ғандарымызға
223	қандарымызда	ғандарымызда	ғандарымызда	ғандарымызда

Продолжение спис. 3

1	2	3	4	5
224	қандарымыздай	ғандарымыздай	ғандарымыздай	ғандарымыздай
225	қандарымыздан	ғандарымыздан	ғандарымыздан	ғандарымыздан
226	қандарымызды	ғандарымызды	ғандарымызды	ғандарымызды
227	қандарымыздың	ғандарымыздың	ғандарымыздың	ғандарымыздың
228	қандарын	ғандарын	ғандарын	ғандарын
229	қандарына	ғандарына	ғандарына	ғандарына
230	қандарынан	ғандарынан	ғандарынан	ғандарынан
231	қандарында	ғандарында	ғандарында	ғандарында
232	қандарындай	ғандарындай	ғандарындай	ғандарындай
233	қандарының	ғандарының	ғандарының	ғандарының
234	қандарың	ғандарың	ғандарың	ғандарың
235	қандарыңа	ғандарыңа	ғандарыңа	ғандарыңа
236	қандарыңда	ғандарыңда	ғандарыңда	ғандарыңда
237	қандарыңдай	ғандарыңдай	ғандарыңдай	ғандарыңдай
238	қандарыңды	ғандарыңды	ғандарыңды	ғандарыңды
239	қандарыңмен	ғандарыңмен	ғандарыңмен	ғандарыңмен
240	қандарыңменен	ғандарыңменен	ғандарыңменен	ғандарыңменен
241	қандарыңнан	ғандарыңнан	ғандарыңнан	ғандарыңнан
242	қандарыңның	ғандарыңның	ғандарыңның	ғандарыңның
243	қандарыңыз	ғандарыңыз	ғандарыңыз	ғандарыңыз
244	қандарыңызбен	ғандарыңызбен	ғандарыңызбен	ғандарыңызбен
245	қандарыңызбенен	ғандарыңызбенен	ғандарыңызбенен	ғандарыңызбенен
246	қандарыңызға	ғандарыңызға	ғандарыңызға	ғандарыңызға
247	қандарыңызда	ғандарыңызда	ғандарыңызда	ғандарыңызда
248	қандарыңыздай	ғандарыңыздай	ғандарыңыздай	ғандарыңыздай
249	қандарыңыздан	ғандарыңыздан	ғандарыңыздан	ғандарыңыздан
250	қандарыңызды	ғандарыңызды	ғандарыңызды	ғандарыңызды
251	қандарыңыздың	ғандарыңыздың	ғандарыңыздың	ғандарыңыздың
252	қанды	ғанды	ғанды	ғанды
253	қандықтан	ғандықтан	ғандықтан	ғандықтан
254	қанмен	ғанмен	ғанмен	ғанмен
255	қанменен	ғанменен	ғанменен	ғанменен
256	қанмын	ғанмын	ғанмын	ғанмын

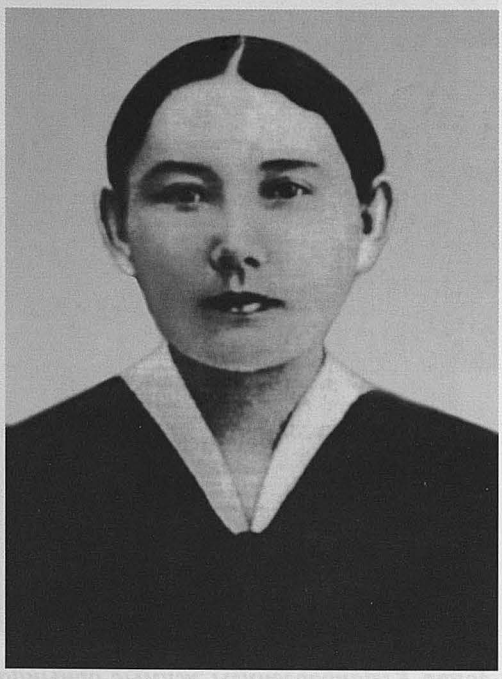
1	2	3	4	5
257	қаннан	ғаннан	ғаннан	ғаннан
258	қанның	ғанның	ғанның	ғанның
259	қансы	ғансы	ғансы	ғансы
260	қансыған	ғансыған	ғансыған	ғансыған
261	қансыған ға	ғансыған ға	ғансыған ға	ғансыған ға
262	қансыған да	ғансыған да	ғансыған да	ғансыған да
263	қансыған дар	ғансыған дар	ғансыған дар	ғансыған дар
264	қансыған дар ға	ғансыған дар ға	ғансыған дар ға	ғансыған дар ға
265	қансыған дар да	ғансыған дар да	ғансыған дар да	ғансыған дар да
266	қансыған дар дан	ғансыған дар дан	ғансыған дар дан	ғансыған дар дан
267	қансыған дар ды	ғансыған дар ды	ғансыған дар ды	ғансыған дар ды
268	қансыған дар дың	ғансыған дар дың	ғансыған дар дың	ғансыған дар дың
269	қансыған дар мен	ғансыған дар мен	ғансыған дар мен	ғансыған дар мен
270	қансыған дар менен	ғансыған дар менен	ғансыған дар менен	ғансыған дар менен
271	қансыған ды	ғансыған ды	ғансыған ды	ғансыған ды
272	қансыған мен	ғансыған мен	ғансыған мен	ғансыған мен
273	қансыған менен	ғансыған менен	ғансыған менен	ғансыған менен
274	қансыған нан	ғансыған нан	ғансыған нан	ғансыған нан
275	қансыған ның	ғансыған ның	ғансыған ның	ғансыған ның
276	қансыз	ғансыз	ғансыз	ғансыз
277	қансың	ғансың	ғансың	ғансың
278	қанша	ғанша	ғанша	ғанша
279	қаным	ғаным	ғаным	ғаным
280	қаным а	ғаным а	ғаным а	ғаным а
281	қаным да	ғаным да	ғаным да	ғаным да
282	қаным дай	ғаным дай	ғаным дай	ғаным дай
283	қаным ды	ғаным ды	ғаным ды	ғаным ды
284	қаным мен	ғаным мен	ғаным мен	ғаным мен
285	қаным менен	ғаным менен	ғаным менен	ғаным менен
286	қаным нан	ғаным нан	ғаным нан	ғаным нан
287	қаным ның	ғаным ның	ғаным ның	ғаным ның
288	қаным ыз	ғаным ыз	ғаным ыз	ғаным ыз
289	қаным ызбен	ғаным ызбен	ғаным ызбен	ғаным ызбен

Продолжение спис. 3

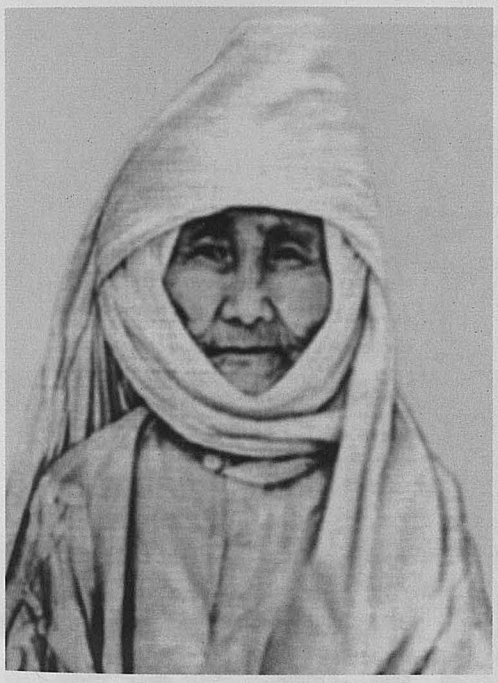
1	2	3	4	5
290	қанымызбенен	ғанымызбенен	ғанымызбенен	ғанымызбенен
291	қанымызға	ғанымызға	ғанымызға	ғанымызға
292	қанымызда	ғанымызда	ғанымызда	ғанымызда
293	қанымыздай	ғанымыздай	ғанымыздай	ғанымыздай
294	қанымыздан	ғанымыздан	ғанымыздан	ғанымыздан
295	қанымызды	ғанымызды	ғанымызды	ғанымызды
296	қанымыздың	ғанымыздың	ғанымыздың	ғанымыздың
297	қаның	ғаның	ғаның	ғаның
298	қаныңа	ғаныңа	ғаныңа	ғаныңа
299	қаныңда	ғаныңда	ғаныңда	ғаныңда
300	қаныңдай	ғаныңдай	ғаныңдай	ғаныңдай
301	қаныңды	ғаныңды	ғаныңды	ғаныңды
302	қаныңмен	ғаныңмен	ғаныңмен	ғаныңмен
303	қаныңменен	ғаныңменен	ғаныңменен	ғаныңменен
304	қаныңнан	ғаныңнан	ғаныңнан	ғаныңнан
305	қаныңның	ғаныңның	ғаныңның	ғаныңның
306	қаныңыз	ғаныңыз	ғаныңыз	ғаныңыз
307	қаныңызбен	ғаныңызбен	ғаныңызбен	ғаныңызбен
308	қаныңызбенен	ғаныңызбенен	ғаныңызбенен	ғаныңызбенен
309	қаныңызға	ғаныңызға	ғаныңызға	ғаныңызға
310	қаныңызда	ғаныңызда	ғаныңызда	ғаныңызда
311	қаныңыздай	ғаныңыздай	ғаныңыздай	ғаныңыздай
312	қаныңыздан	ғаныңыздан	ғаныңыздан	ғаныңыздан
313	қаныңызды	ғаныңызды	ғаныңызды	ғаныңызды
314	қаныңыздың	ғаныңыздың	ғаныңыздың	ғаныңыздың
315	қылары	ғылары	ғылары	ғылары
316	қыларымыз	ғыларымыз	ғыларымыз	ғыларымыз
317	қыларың	ғыларың	ғыларың	ғыларың
318	қыларыңыз	ғыларыңыз	ғыларыңыз	ғыларыңыз
319	қым	ғым	ғым	ғым
320	қымыз	ғымыз	ғымыз	ғымыз
321	қың	ғың	ғың	ғың
322	қыңыз	ғыңыз	ғыңыз	ғыңыз



Асқардың әкесі
Құдайберген Қуанұлы Жұбанов



Асқардың анасы
Раушан Оспанқызы Жұбанова



Асқардың әжесі
Бибішынар



Құдайбергеннің інісі
Ахмет Жұбанов



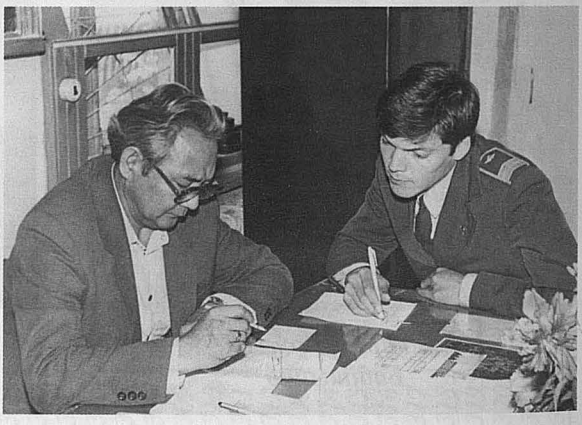
Асқар Құдайбергенұлы жұмыс орнында



А. Жұбановтың әскери борышын өтеп жүрген кезі



Асқар Құдайбергенұлы Қызғалдақ және Қырмызы апаларымен бірге. 1950 ж.



А.Қ. Жұбанов "жоғары математика" пәнінің мұғалімі



Қ.Қ. Жұбанов отбасында.
Алдыңғы қатарда сол жақтан А. Жұбанов, Қ. Жұбанов, ғалымның қызы Мүсілима, жұбайы Раушан мен ұлы Ақырап. Екінші қатарда педтехникум студенттері Х. Әуезов пен ғалымның кіші інісі М. Жұбанов. Ақтөбе, 1928 ж.



Қ. Жұбановтың туғанына 65 жыл. Алдыңғы қатарда академиктер М.С. Сильченко, А.Қ. Жұбанов, профессор М. Балақаев. Екінші қатарда ғалымның балалары: Ақырап, Қызғалдақ, Мүсілима, Асқар. Одан кейін ғалымның шәкірті М. Қаратаев, тілші-ғалым Р. Сыздық және Есет Құдайбергенұлы.
Алматы, 1964 ж.



Қ. Жұбановтың 100 жылдығына арналып өткізілген ғылыми конференция кезі. Суретте ғалымның балалары мен туыстары



Конференция қонақтары ғалым атындағы университеттің Жұбановтану зертханасында. Ақтөбе



Мәскеу ғалымы Р.К. Потаповамен кездесу



Аспирант А.Қ. Жұбанов Минск-22 ЭЕМ-ның басқару жүйесі алдында





Профессор Асқар Құдайбергенұлы Жұбанов жетекшілік ететін "компьютерлік лингвистика" зертханасында



Әке ескерткішіне ұрпақтар тағзымы. Ақтөбе

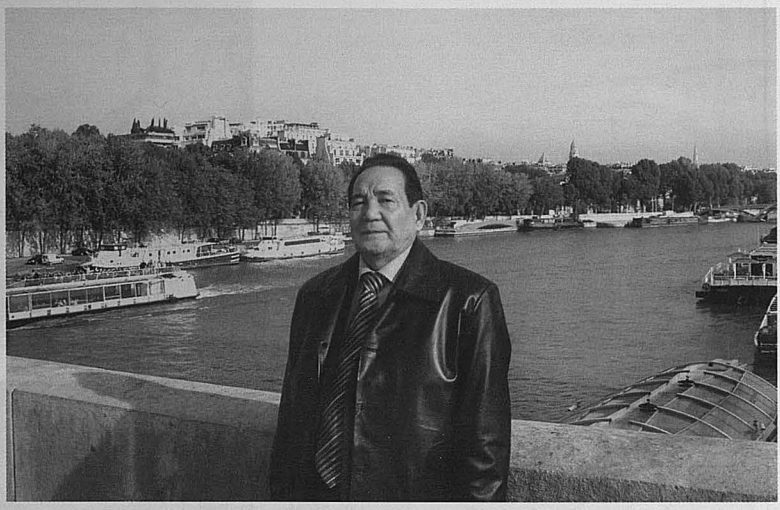


Ақтөбедегі Қ. Жұбанов ескерткішіне Елбасымыз Н.Ә. Назарбаевтың гүлшоғын қою сәті



Қ. Жұбановтың 80 жылдығына арналған көрмеден көрініс. Ортада ғалымның жұбайы Раушан, оң жақта оның ортаншы қызы Қызғалдақ

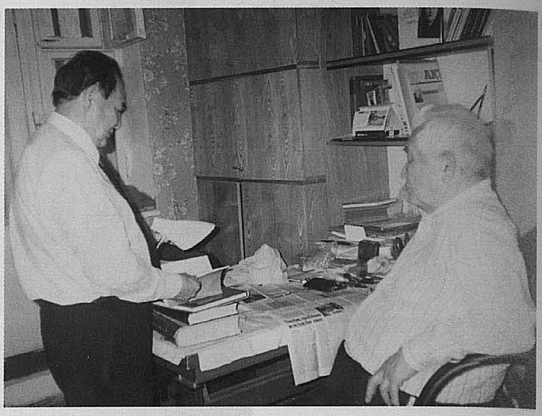
Халықаралық ЮНЕСКО ұйымының қарарымен Ахмет Жұбановтың 100 жылдығына арналып Парижде өткізілген ғылым мен мәдениет күндерінде. 12-17 қазан, 2006 ж.





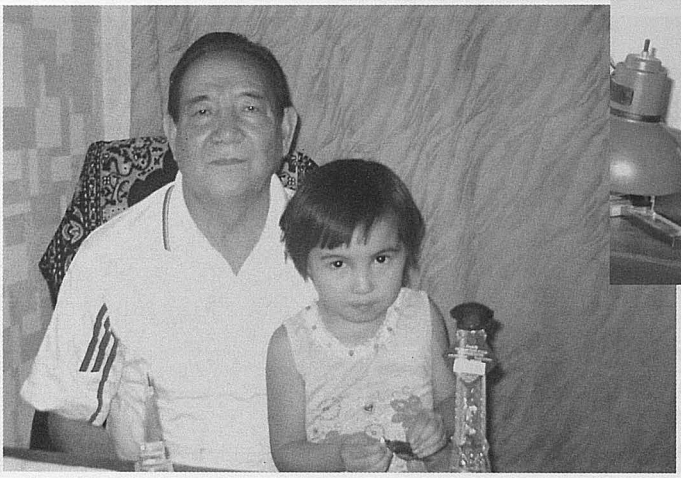


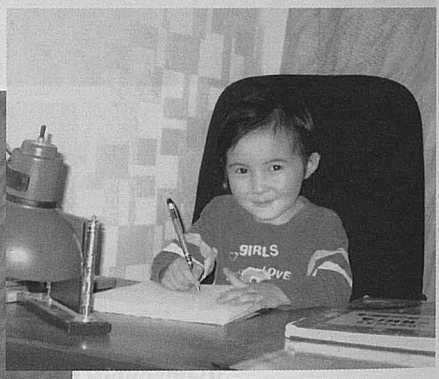
“Кеңесіп пішкен тон келте болмас”. Ағайынды Асқар мен Есет





А. Байтұрсынұлы атындағы Тіл білімі институты ғалымдарының демалыс сәтінен көрініс





Ғалым қызы Еңлікпен бірге. 2007 ж.

1	2	3	4	5
323	қысы	ғысы	ғысы	ғысы
324	па	ма	ба	ма
235	паган	маган	баган	маган
326	пагапга	маганга	баганга	маганга
327	паганда	маганда	баганда	маганда
328	пагандай	магандай	багандай	магандай
329	пагандар	магандар	багандар	магандар
330	пагандарға	магандарға	багандарға	магандарға
331	пагандарда	магандарда	багандарда	магандарда
332	пагандардан	магандардан	багандардан	магандардан
333	пагандарды	магандарды	багандарды	магандарды
334	пагандардың	магандардың	багандардың	магандардың
335	пагандармен	магандармен	багандармен	магандармен
336	пагандарменен	магандарменен	багандарменен	магандарменен
337	паганды	маганды	баганды	маганды
338	пагандықтан	магандықтан	багандықтан	магандықтан
339	паганмен	маганмен	баганмен	маганмен
340	паганменен	маганменен	баганменен	маганменен
341	паганнан	маганнан	баганнан	маганнан
342	паганның	маганның	баганның	маганның
343	пагансы	магансы	багансы	магансы
344	паганша	маганша	баганша	маганша
345	пагы	магы	багы	магы
346	пагым	магым	багым	магым
347	пагыма	магыма	багыма	магыма
348	пагымда	магымда	багымда	магымда
349	пагымды	магымды	багымды	магымды
350	пагымен	магымен	багымен	магымен
351	пагыменен	магыменен	багыменен	магыменен
352	пагыммен	магыммен	багыммен	магыммен
353	пагымменен	магымменен	багымменен	магымменен
354	пагымнан	магымнан	багымнан	магымнан
355	пагымның	магымның	багымның	магымның

Продолжение спис. 3

1	2	3	4	5
356	пагымыз	магымыз	багымыз	магымыз
357	пагымызбен	магымызбен	багымызбен	магымызбен
358	пагымызбенен	магымызбенен	багымызбенен	магымызбенен
359	пагымызга	магымызга	багымызга	магымызга
360	пагымызда	магымызда	багымызда	магымызда
361	пагымыздан	магымыздан	багымыздан	магымыздан
362	пагымызды	магымызды	багымызды	магымызды
363	пагымыздың	магымыздың	багымыздың	магымыздың
364	пагына	магына	багына	магына
365	пагынан	магынан	багынан	магынан
366	пагында	магында	багында	магында
367	пагының	магының	багының	магының
368	пагың	магың	багың	магың
369	пагыңа	магыңа	багыңа	магыңа
370	пагыңда	магыңда	багыңда	магыңда
371	пагыңды	магыңды	багыңды	магыңды
372	пагыңмен	магыңмен	багыңмен	магыңмен
373	пагыңменен	магыңменен	багыңменен	магыңменен
374	пагыңнан	магыңнан	багыңнан	магыңнан
375	пагыңның	магыңның	багыңның	магыңның
376	пагыңыз	магыңыз	багыңыз	магыңыз
377	пагыңызбен	магыңызбен	багыңызбен	магыңызбен
378	пагыңызбенен	магыңызбенен	багыңызбенен	магыңызбенен
379	пагыңызга	магыңызга	багыңызга	магыңызга
380	пагыңызда	магыңызда	багыңызда	магыңызда
381	пагыңыздан	магыңыздан	багыңыздан	магыңыздан
382	пагыңызды	магыңызды	багыңызды	магыңызды
383	пагыңыздың	магыңыздың	багыңыздың	магыңыздың
384	пайды	майды	байды	майды
385	паймыз	маймыз	баймыз	маймыз
386	паймын	маймын	баймын	маймын
387	пайсың	майсың	байсың	майсың
388	пайсыңдар	майсыңдар	байсыңдар	майсыңдар

Продолжение спис. 3

1	2	3	4	5
389	*пайтын*	*майтын*	*байтын*	*майтын*
390	*пайтынға*	*майтынға*	*байтынға*	*майтынға*
391	*пайтында*	*майтында*	*байтында*	*майтында*
392	*пайтындар*	*майтындар*	*байтындар*	*майтындар*
393	*пайтындарға*	*майтындарға*	*байтындарға*	*майтындарға*
394	*пайтындарда*	*майтындарда*	*байтындарда*	*майтындарда*
395	*пайтындардан*	*майтындардан*	*байтындардан*	*майтындардан*
396	*пайтындарды*	*майтындарды*	*байтындарды*	*майтындарды*
397	*пайтындардың*	*майтындардың*	*байтындардың*	*майтындардың*
398	*пайтындармен*	*майтындармен*	*байтындармен*	*майтындармен*
399	*пайтындарменен*	*майтындарменен*	*байтындарменен*	*майтындарменен*
400	*пайтынды*	*майтынды*	*байтынды*	*майтынды*
401	*пайтынмен*	*майтынмен*	*байтынмен*	*майтынмен*
402	*пайтынменен*	*майтынменен*	*байтынменен*	*майтынменен*
403	*пайтыннан*	*майтыннан*	*байтыннан*	*майтыннан*
404	*пайтынның*	*майтынның*	*байтынның*	*майтынның*
405	*пайық*	*майық*	*байық*	*майық*
406	*пайықшы*	*майықшы*	*байықшы*	*майықшы*
407	*пақ*	*мақ*	*бақ*	*мақ*
408	*паққа*	*мақ̆қа*	*баққа*	*мақ̆қа*
409	*пақпен*	*мақпен*	*бақпен*	*мақпен*
410	*пақпенен*	*мақпенен*	*бақпенен*	*мақпенен*
411	*пақпын*	*мақпын*	*бақпын*	*мақпын*
412	*пақсыз*	*мақсыз*	*бақсыз*	*мақсыз*
413	*пақсыздар*	*мақсыздар*	*бақсыздар*	*мақсыздар*
414	*пақсың*	*мақсың*	*бақсың*	*мақсың*
415	*пақсыңдар*	*мақсыңдар*	*бақсыңдар*	*мақсыңдар*
416	*пақта*	*мақта*	*бақта*	*мақта*
417	*пақтай*	*мақтай*	*бақтай*	*мақтай*
418	*пақтан*	*мақтан*	*бақтан*	*мақтан*
419	*пақты*	*мақты*	*бақты*	*мақты*
420	*пақтың*	*мақтың*	*бақтың*	*мақтың*
421	*пақшы*	*мақшы*	*бақшы*	*мақшы*

Продолжение спис. 3

1	2	3	4	5
422	паңыз	маңыз	баңыз	маңыз
423	паңыздар	маңыздар	баңыздар	маңыздар
424	паңыздаршы	маңыздаршы	баңыздаршы	маңыздаршы
425	пас	мас	бас	мас
426	паса	маса	баса	маса
427	пасаң	масаң	басаң	масаң
428	пасаңдар	масаңдар	басаңдар	масаңдар
429	пасаңдаршы	масаңдаршы	басаңдаршы	масаңдаршы
430	пасаңшы	масаңшы	басаңшы	масаңшы
431	пасаңыз	масаңыз	басаңыз	масаңыз
432	пасаңыздар	масаңыздар	басаңыздар	масаңыздар
433	пасаңыздаршы	масаңыздаршы	басаңыздаршы	масаңыздаршы
434	пасаңызшы	масаңызшы	басаңызшы	масаңызшы
435	пасбыз	масбыз	басбыз	масбыз
436	пасқа	масқа	басқа	масқа
437	паспен	маспен	баспен	маспен
438	паспенен	маспенен	баспенен	маспенен
439	пассыз	массыз	бассыз	массыз
440	пассыздар	массыздар	бассыздар	массыздар
441	пассың	массың	бассың	массың
442	пассыңдар	массыңдар	бассыңдар	массыңдар
443	паста	маста	баста	маста
444	пастай	мастай	бастай	мастай
445	пастан	мастан	бастан	мастан
446	пастар	мастар	бастар	мастар
447	пастарға	мастарға	бастарға	мастарға
448	пастарда	мастарда	бастарда	мастарда
449	пастардан	мастардан	бастардан	мастардан
450	пастарды	мастарды	бастарды	мастарды
451	пастардың	мастардың	бастардың	мастардың
452	пастармен	мастармен	бастармен	мастармен
453	пастарменен	мастарменен	бастарменен	мастарменен
454	пасты	масты	басты	масты

Продолжение спис. 3

1	2	3	4	5
455	пастың	мастың	бастың	мастың
456	пасы	масы	басы	масы
457	пасым	масым	басым	масым
458	пасыма	масыма	басыма	масыма
459	пасымда	масымда	басымда	масымда
460	пасымды	масымды	басымды	масымды
461	пасымен	масымен	басымен	масымен
462	пасыменен	масыменен	басыменен	масыменен
463	пасыммен	масыммен	басыммен	масыммен
464	пасымменен	масымменен	басымменен	масымменен
465	пасымнан	асымнан	басымнан	масымнан
466	пасымның	масымның	басымның	масымның
467	пасымыз	масымыз	басымыз	масымыз
468	пасымызбен	масымызбен	басымызбен	масымызбен
469	пасымызбенен	масымызбенен	басымызбенен	масымызбенен
470	пасымызга	масымызга	басымызга	масымызга
471	пасымызда	масымызда	басымызда	масымызда
472	пасымыздан	масымыздан	басымыздан	масымыздан
473	пасымызды	масымызды	басымызды	масымызды
474	пасымыздың	масымыздың	басымыздың	масымыздың
475	пасына	масына	басына	масына
476	пасынан	масынан	басынан	масынан
477	пасының	масының	басының	масының
478	пасың	масың	басың	масың
479	пасыңа	масыңа	басыңа	масыңа
480	пасыңда	масыңда	басыңда	масыңда
481	пасыңды	масыңды	басыңды	масыңды
482	пасыңмен	масыңмен	басыңмен	масыңмен
483	пасыңменен	масыңменен	басыңменен	масыңменен
484	пасыңнан	масыңнан	басыңнан	масыңнан
485	пасыңның	масыңның	басыңның	масыңның
486	пасыңыз	масыңыз	басыңыз	масыңыз
487	пасыңызбен	масыңызбен	басыңызбен	масыңызбен

Продолжение спис. 3

1	2	3	4	5
488	пасыңызбенен	масыңызбенен	басыңызбенен	масыңызбенен
489	пасыңызга	масыңызга	басыңызга	масыңызга
490	пасыңызда	масыңызда	басыңызда	масыңызда
491	пасыңыздан	масыңыздан	басыңыздан	масыңыздан
492	пасыңызды	масыңызды	басыңызды	масыңызды
493	пасыңыздың	масыңыздың	басыңыздың	масыңыздың
494	пау	мау	бау	мау
495	пауга	мауга	бауга	мауга
496	пауда	мауда	бауда	мауда
497	паудан	маудан	баудан	маудан
498	пауды	мауды	бауды	мауды
499	паудың	маудың	баудың	маудың
500	паулары	маулары	баулары	маулары
501	пауларымен	мауларымен	бауларымен	мауларымен
502	пауларыменен	мауларыменен	бауларыменен	мауларыменен
503	пауларымыз	мауларымыз	бауларымыз	мауларымыз
504	пауларымызбен	мауларымызбен	бауларымызбен	мауларымызбен
505	пауларымызбенен	мауларымызбенен	бауларымызбенен	мауларымызбенен
506	пауларымызга	мауларымызга	бауларымызга	мауларымызга
507	пауларымызда	мауларымызда	бауларымызда	мауларымызда
508	пауларымыздан	мауларымыздан	бауларымыздан	мауларымыздан
509	пауларымызды	мауларымызды	бауларымызды	мауларымызды
510	пауларымыздың	мауларымыздың	бауларымыздың	мауларымыздың
511	пауларына	мауларына	бауларына	мауларына
512	пауларынан	мауларынан	бауларынан	мауларынан
513	пауларында	мауларында	бауларында	мауларында
514	пауларының	мауларының	бауларының	мауларының
515	пауларың	мауларың	бауларың	мауларың
516	пауларыңа	мауларыңа	бауларыңа	мауларыңа
517	пауларыңда	мауларыңда	бауларыңда	мауларыңда
518	пауларыңды	мауларыңды	бауларыңды	мауларыңды
519	пауларыңмен	мауларыңмен	бауларыңмен	мауларыңмен
520	пауларыңменен	мауларыңменен	бауларыңменен	мауларыңменен

Продолжение спис. 3

1	2	3	4	5
521	пауларыңнан	мауларыңнан	бауларыңнан	мауларыңнан
522	пауларыңның	мауларыңның	бауларыңның	мауларыңның
523	пауларыңыз	мауларыңыз	бауларыңыз	мауларыңыз
524	пауларыңызбен	мауларыңызбен	бауларыңызбен	мауларыңызбен
525	пауларыңызбенен	мауларыңызбенен	бауларыңызбенен	мауларыңызбенен
526	пауларыңызға	мауларыңызға	бауларыңызға	мауларыңызға
527	пауларыңызда	мауларыңызда	бауларыңызда	мауларыңызда
528	пауларыңыздан	мауларыңыздан	бауларыңыздан	мауларыңыздан
529	пауларыңызды	мауларыңызды	бауларыңызды	мауларыңызды
530	пауларыңыздың	мауларыңыздың	бауларыңыздың	мауларыңыздың
531	паумен	маумен	баумен	маумен
532	пауменен	мауменен	бауменен	мауменен
533	паушы	маушы	баушы	маушы
534	пауы	мауы	бауы	мауы
535	пауым	мауым	бауым	мауым
536	пауыма	мауыма	бауыма	мауыма
537	пауымда	мауымда	бауымда	мауымда
538	пауымды	мауымды	бауымды	мауымды
539	пауымен	мауымен	бауымен	мауымен
540	пауыменен	мауыменен	бауыменен	мауыменен
541	пауыммен	мауыммен	бауыммен	мауыммен
542	пауымменен	мауымменен	бауымменен	мауымменен
543	пауымнан	мауымнан	бауымнан	мауымнан
544	пауымның	мауымның	бауымның	мауымның
545	пауымыз	мауымыз	бауымыз	мауымыз
546	пауымызбен	мауымызбен	бауымызбен	мауымызбен
547	пауымызбенен	мауымызбенен	бауымызбенен	мауымызбенен
548	пауымызға	мауымызға	бауымызға	мауымызға
549	пауымызда	мауымызда	бауымызда	мауымызда
550	пауымыздан	мауымыздан	бауымыздан	мауымыздан
551	пауымызды	мауымызды	бауымызды	мауымызды
552	пауымыздың	мауымыздың	бауымыздың	мауымыздың
553	пауына	мауына	бауына	мауына

Продолжение спис. 3

1	2	3	4	5
554	пауынан	мауынан	бауынан	мауынан
555	пауының	мауының	бауының	мауының
556	пауың	мауың	бауың	мауың
557	пауыңа	мауыңа	бауыңа	мауыңа
558	пауыңда	мауыңда	бауыңда	мауыңда
559	пауыңды	мауыңды	бауыңды	мауыңды
560	пауыңмен	мауыңмен	бауыңмен	мауыңмен
561	пауыңменен	мауыңменен	бауыңменен	мауыңменен
562	пауыңнан	мауыңнан	бауыңнан	мауыңнан
563	пауыңның	мауыңның	бауыңның	мауыңның
564	пауыңыз	мауыңыз	бауыңыз	мауыңыз
565	пауыңызбен	мауыңызбен	бауыңызбен	мауыңызбен
566	пауыңызбенен	мауыңызбенен	бауыңызбенен	мауыңызбенен
567	пауыңызга	мауыңызга	бауыңызга	мауыңызга
568	пауыңызда	мауыңызда	бауыңызда	мауыңызда
569	пауыңыздан	мауыңыздан	бауыңыздан	мауыңыздан
570	пауыңызды	мауыңызды	бауыңызды	мауыңызды
571	пауыңыздың	мауыңыздың	бауыңыздың	мауыңыздың
572	са	са	са	са
573	сам	сам	сам	сам
574	самшы	самшы	самшы	самшы
575	саң	саң	саң	саң
576	саңдар	саңдар	саңдар	саңдар
577	саңдаршы	саңдаршы	саңдаршы	саңдаршы
578	саңшы	саңшы	саңшы	саңшы
579	саңыз	саңыз	саңыз	саңыз
580	саңыздар	саңыздар	саңыздар	саңыздар
581	саңыздаршы	саңыздаршы	саңыздаршы	саңыздаршы
582	саңызшы	саңызшы	саңызшы	саңызшы
583	сын	сын	сын	сын
584	сыншы	сыншы	сыншы	сыншы
585	ты	ды	ды	ды
586	тым	дым	дым	дым

1	2	3	4	5
587	тың	дың	дың	дың
588	тыңдар	дыңдар	дыңдар	дыңдар
589	у	у	у	у
590	уга	уга	уга	уга
591	уда	уда	уда	уда
592	удан	удан	удан	удан
593	уды	уды	уды	уды
594	удың	удың	удың	удың
595	улары	улары	улары	улары
596	уларым	уларым	уларым	уларым
597	уларыма	уларыма	уларыма	уларыма
598	уларымда	уларымда	уларымда	уларымда
599	уларымды	уларымды	уларымды	уларымды
600	уларымен	уларымен	уларымен	уларымен
601	уларыменен	уларыменен	уларыменен	уларыменен
602	уларыммен	уларыммен	уларыммен	уларыммен
603	уларымменен	уларымменен	уларымменен	уларымменен
604	уларымнан	уларымнан	уларымнан	уларымнан
605	уларымның	уларымның	уларымның	уларымның
606	уларымыз	уларымыз	уларымыз	уларымыз
607	уларымызбен	уларымызбен	уларымызбен	уларымызбен
608	уларымызбенен	уларымызбенен	уларымызбенен	уларымызбенен
609	уларымызга	уларымызга	уларымызга	уларымызга
610	уларымызда	уларымызда	уларымызда	уларымызда
611	уларымыздан	уларымыздан	уларымыздан	уларымыздан
612	уларымызды	уларымызды	уларымызды	уларымызды
613	уларымыздың	уларымыздың	уларымыздың	уларымыздың
614	уларына	уларына	уларына	уларына
615	уларынан	уларынан	уларынан	уларынан
616	уларының	уларының	уларының	уларының
617	уларың	уларың	уларың	уларың
618	уларыңа	уларыңа	уларыңа	уларыңа
619	уларыңда	уларыңда	уларыңда	уларыңда

Продолжение спис. 3

1	2	3	4	5
620	уларыңды	уларыңды	уларыңды	уларыңды
621	уларыңмен	уларыңмен	уларыңмен	уларыңмен
622	уларыңменен	уларыңменен	уларыңменен	уларыңменен
623	уларыңнан	уларыңнан	уларыңнан	уларыңнан
624	уларыңның	уларыңның	уларыңның	уларыңның
625	уларыңыз	уларыңыз	уларыңыз	уларыңыз
626	уларыңызбен	уларыңызбен	уларыңызбен	уларыңызбен
627	уларыңызбенен	уларыңызбенен	уларыңызбенен	уларыңызбенен
628	уларыңызга	уларыңызга	уларыңызга	уларыңызга
629	уларыңызда	уларыңызда	уларыңызда	уларыңызда
630	уларыңыздан	уларыңыздан	уларыңыздан	уларыңыздан
631	уларыңызды	уларыңызды	уларыңызды	уларыңызды
632	уларыңыздың	уларыңыздың	уларыңыздың	уларыңыздың
633	умен	умен	умен	умен
634	уменен	уменен	уменен	уменен
635	ушы	ушы	ушы	ушы
636	уы	уы	уы	уы
637	уым	уым	уым	уым
638	уыма	уыма	уыма	уыма
639	уымда	уымда	уымда	уымда
640	уымдай	уымдай	уымдай	уымдай
641	уымды	уымды	уымды	уымды
642	уымен	уымен	уымен	уымен
643	уыменен	уыменен	уыменен	уыменен
644	уыммен	уыммен	уыммен	уыммен
645	уымменен	уымменен	уымменен	уымменен
646	уымнан	уымнан	уымнан	уымнан
647	уымның	уымның	уымның	уымның
648	уымыз	уымыз	уымыз	уымыз
649	уымызбен	уымызбен	уымызбен	уымызбен
650	уымызбенен	уымызбенен	уымызбенен	уымызбенен
651	уымызга	уымызга	уымызга	уымызга
652	уымызда	уымызда	уымызда	уымызда

1	2	3	4	5
653	уымыздай	уымыздай	уымыздай	уымыздай
654	уымыздан	уымыздан	уымыздан	уымыздан
655	уымызды	уымызды	уымызды	уымызды
656	уымыздың	уымыздың	уымыздың	уымыздың
657	уына	уына	уына	уына
658	уынан	уынан	уынан	уынан
659	уында	уында	уында	уында
660	уының	уының	уының	уының
661	уың	уың	уың	уың
662	уыңа	уыңа	уыңа	уыңа
663	уыңда	уыңда	уыңда	уыңда
664	уыңдай	уыңдай	уыңдай	уыңдай
665	уыңды	уыңды	уыңды	уыңды
666	уыңмен	уыңмен	уыңмен	уыңмен
667	уыңменен	уыңменен	уыңменен	уыңменен
668	уыңнан	уыңнан	уыңнан	уыңнан
669	уыңның	уыңның	уыңның	уыңның
670	уыңыз	уыңыз	уыңыз	уыңыз
671	уыңызбен	уыңызбен	уыңызбен	уыңызбен
672	уыңызбенен	уыңызбенен	уыңызбенен	уыңызбенен
673	уыңызга	уыңызга	уыңызга	уыңызга
674	уыңызда	уыңызда	уыңызда	уыңызда
675	уыңыздай	уыңыздай	уыңыздай	уыңыздай
676	уыңыздан	уыңыздан	уыңыздан	уыңыздан
677	уыңызды	уыңызды	уыңызды	уыңызды
678	уыңыздың	уыңыздың	уыңыздың	уыңыздың
679	шы	шы	шы	шы
680	ың	ың	ың	ң
681	ыңдар	ыңдар	ыңдар	ңдар
682	ыңдаршы	ыңдаршы	ыңдаршы	ңдаршы
683	ыңыз	ыңыз	ыңыз	ңыз
684	ыңыздар	ыңыздар	ыңыздар	ңыздар
685	ыңыздаршы	ыңыздаршы	ыңыздаршы	ңыздаршы

Окончание спис. 3

1	2	3	4	5
686	*ыңызшы*	*ыңызшы*	*ыңызшы*	*ңызшы*
687	*ып*	*ып*	*ып*	*п*
688	*ыппыз*	*ыппыз*	*ыппыз*	*ппыз*
689	*ыппын*	*ыппын*	*ыппын*	*ппын*
690	*ыпсыз*	*ыпсыз*	*ыпсыз*	*псыз*
691	*ыпсыздар*	*ыпсыздар*	*ыпсыздар*	*псыздар*
692	*ыпсың*	*ыпсың*	*ыпсың*	*псың*
693	*ыпсыңдар*	*ыпсыңдар*	*ыпсыңдар*	*псыңдар*
694	*ыпты*	*ыпты*	*ыпты*	*пты*

Фрагмент 6

В пятой главе книги на основе осмысления содержаний и анализа статистических данных следующих минитекстов были составлены: 1) XUD-03: *табл. 26* — фрагмент потенциального алфавитно-частотного распределительного словаря, *табл.28* — таблица основного статического содержания (ТОСС); 2) PUB-06 *табл. 29* — таблица основного статического содержания; 3) NAU-15 *табл. 30* — таблица основного статического содержания.

Ниже приводим эти минитексты, относящиеся к трем функционально-смысловым типам как повествование, описание и рассуждение:

XUD-03:

Сейсенбаев ояздық атқару комитетінің нұсқаушысы еді. Ояздық атқару комитеттен Борлыкөл елінің болыстық атқару комитетінің бір шатақ жұмыстарына барған. Қайтып келе жатып, Шиелі поселкесінің ар жағында 10–15 шақырымдай жердегі, бұрын би болған, белгілі, ақсақ Ақан дегендікіне қоныпты. Өңірде ас пісіп, Сейсенбаев, ақсақ Ақан және үш-төрт ауыл кісілері табақты жаба алдарына алып, еттің буын бұрқыратып турай бергенде, үйге сол ауданда жүрген белгілі банды Кудря кіріп келіпті. Кудря жұрттың айтуынша және істеп жүргендеріне қарағанда, өзі балуан, жасаулы мылтықтан тайсалу дегенді білмейтін, талайларды өлтірген. Міне, сол Кудря қасында ешкім жоқ, алтыатарын қолына ұстап, үйге кіріп келіп тұра қалыпты:

— Кәне, мынау отырған исполкомның нұсқаушысы болса, тысқа шықсын! — депті.

Сейсенбаев жылап қоя беріп, ақсақ Ақанның артына тығылып, құшақтап жалыныпты. Сол жерде ақсақ Ақан араға түсіп, Кудряны тоқтатып, Сейсенбаевты аман алып қалыпты. Сейсенбаевқа білдірмей, біреу политбюроға мұны құпиялап баяндапты...

Міне, енді, Кудряның ізін, кімдермен байланыс қылатынын астыртын жансыз жүріп біліп, ұстауға әрекет қылуға политбюро Хамитті жұмсап отыр. Кудряның шарлап жүрген ізінің ұшығын қазақ арасынан оңайырақ табуға мүмкін гой деп ойласып, Хамиттің баруын мақұлдады.

Хамит кешікпей керек-жарақ саймандарының бәрін түгелдеп алып, бес орыс жолдаспен жүріп кетті.

Ақмолаға баратын үлкен жолмен жүріп келіп, Шиелі поселкесіне таянғанда Хамит жолдастарымен жүріс-тұрыстың, байланысып, хабарласып отырудың жобасын әбден ақылдасып, жасасып алып, жолдастарын қоя беріп, өзі бұрылып қазақ аулына кетті.

Хамит — ұзынша бойлы, жауырынды, тіп-тік, сұр жігіт. Қырша мұрынды, шүңіректеу қара кер тобылғы көзді, қырқып жүретін сақалы, мұрты бар. Жасы 25-26 шамасында. Сом, сұлу денесі ылғи сіңірмен шиырылған күшті көк ет еді. Бар денесі серіппелі көк темірден құйып жасағандай. Түйіліп жазылған қиғаштау қасы мен шұңғылдау көзі Хамитте зор қайрат, үлкен қайсарлық бар екенін көрсетеді. Бет-аузы бүркіттің бетіндей. Қыран бүркіттің тұрпілі саусақтарындай, күшті темірдей жұдырықтарын түйіп, көрін тігін, кіжініп:

— А — а... қандай қылайын, — деп төніп келсе, жөпшенді кісінің аруағы ұшып кететіндей азамат. Хамит салт атты. Үстінде мұздай қазақша киім. Басында қоңыр елтірі тымақ. Аты сұлы жеп жүрген, мейіздей қатқан, сұлу Құбақан ат.

Ол күні қазақ аулына қонып, елеусіз отырып, сөзден сөз туғызып бандылардың жайынан әңгіме қылса да, қонған үйдің иесінен де және ауыл жігіттерінен де мандымды сөзге жарымай ілгері жүріп кетті.

Ақсақ Ақан аулына барып қонды. Одан біраз сөзге қанып аттанып, бандының ар жақ беттегі Ақмоламен шектес елдер жаққа кеткен дыбысы сезіліп, солай жүрді.

PUB-06: ***ПӘН АПТАЛЫҒЫ***

Ана тілі мен әдебиет секциясының ұйымдастыруымен О.Жандосов атындағы орта мектепте пән апталығы өтті. Онда алғашқы күні пән секциясының мүшелері өз сабақтарында ана тілінің қоғам мен жеке адам өмірінде алатын орны мен маңызын түсіндірді. Тіл мен әдебиет туралы оқушыларға қызықты әңгімелер айтты. "Ана тілім — асыл қазынам" атты қабырға газеті шығарылды.

Екінші күн пән мұғалімдері оқушылардың ана тілі мен әдебиетінен алған білімдерін тереңдету, өздігінен жұмыс істеу дағдыларын жетілдіру мақсатында Отан, партия, бейбітшілік тақырыбында жаттауға өлеңдер берді. Баяндама, еркін тақырыпқа шығарма жазуды тапсырды. Дидактикалық материалдар, көрнекі құралдар жасатты.

Апталықтың үшінші күні қазақ тілі мен әдебиет пәнінің мұғалімі М.Кенжегұлов ашық сабақ өткізді. Мектеп кітапханашысы К.Нұрпейісова 5-6 класс оқушыларымен балалар жазушысы С.Бегалиннің туғанына 90 жыл толуына орай оның "Сәтжан" повесі бойынша оқырмандар конференциясын өткізді.

Төртінші күні апталықтың қорытынды кешінің жоспары жасалды, оқушылыр оған дайындалды.

Апталықтың бесінші күні 7 "А" класында "Ғылым таппай мақтанба" деген тақырыпта ұлы ақын Абай Құнанбаевтың творчествосына арналған ашық тәрбие сағаты өтті.

Сенбі күні мектептің акт залында оқушылар творчествосына арналған көрме, кітап бұрышы ұйымдастырылды. "Ана тілім — асыл қазынам" атты пән апталығының қорытынды кеші тартымды өтті. Онда шәкірттер "Мектеп — кеме, білім — теңіз", "Тыныс белгілері" атты әдеби-монтаж ұйымдастырды.

NAU-15: ***Минералдар құрылысы***

Геология музейінен көптеген минералдардың әдемі кристалдарын көруге болады. Олардың неше түрлі қырлы кристалдары өздерінің формаларының әдемілігімен көз тұндырады. Осыдан келіп барлық минералдар ірі кристалл түзе ме деген сұрақ туады. Онда біз оларды неге сирек кездестіреміз? Міне, сондықтан да біз енді минералдардың құрылысына назар аударамақпыз. Ал минералдардың құрылысындағы ерекшеліктердің флотация үшін маңызы зор.

Заттарды құрайтын ұсақ бөлшектер — атомдар, иондар және молекулалар қандай да бір заттың құрамында өзара белгілі бір тәртіппен кристалдар түзеді. Кристалл газдардан да, балқыған заттардан да және ерітінділерден де пайда болады. Кейде тіпті кристалл емес кейбір қатты заттар (аморфты заттар) белгілі бір жағдайлардың әсерінен біртіндеп кристалға айналады. Оған мысал ретінде, шыныны алуға болады.

Бізді өзінің әдемілігімен таңдандыратын ірі кристалдар көбінесе ерітінділерден пайда болады. Ал газдардан, балқыған заттардан түзілетін кристалдар көбінесе ұсақ, кейде тіпті жай көзге көрінбейтін майда түрде түзіледі. Ерітінділерден де үнемі ірі кристалдар пайда бола бермейді. Сондықтан көптеген минералдардың ірі кристалдары жиі кездесе бермейді. Ал оларды арнаулы микроскоппен қарағанда барлық қатты заттардың кристалдардан тұратынын анықтауға болады.

Сонда ірі кристалдар қалай пайда болады? Біріншден, сол затты құрайтын бөлшектер жеткілікті мөлшерде болуы керек. Мысалы, ерітінділер қанық түрде болуы қажет. Екіншіден, кристалдың өсуіне әсер ететін жағдайлар — температура мен қысым. Температура мен қысым неғұрлым ұзақ уақыт өзгермей тұра берсе, кристалл соғұрлым ірілене, яғни өсе береді. Үшіншіден, ерітіндіде, бөтен бөлшектер аз болса, кристалл соғұрлым таза болады. Ал егер кристалдың құрамында бөтен бөлшектер болса, онда оның түсі өзгеріп, әдемілене түседі.

Табиғатта минералдардың ірі кристалдары өте сирек кездеседі. Ең ірі кристалл Бразилиядан табылған. Оның салмағы 40 тоннадай. Тау хрусталінің кісі бойындай ірі кристалдары кездесіп қалады. Олардың арасында ірілігі жағынан ең жиі кездесетіні кварц кристалы. Оның басты себебі сол кварцты (SiO_2) түзетін кремний мен (S) оттегінің (O) жер қыртысында ең көп тараған элемент болуында.

Әрі кристалдардың ерітінділерінде түзілуінің де себептері бар. Ең басты себеп — бөлшектердің су ерітінділерінде қозғалу жылдамдығының балқыған тығыз зат ішіндегі жылдамдықтан жоғарылығында. Ал газдарда бөлшектердің ара қашықтықтары алыс және қозғалыс жылдамдықтары өте жоғары. Соның салдарынан олар жиналып кристалдана алмайды. Мысалы, атмосферадағы су молекулалары қысты күні ауаның температурасы төмендегенде қозғалыс жылдамдығының азаюы салдарынан бірігіп, қарға айналып, бәрімізге белгілі бес бұрышты кристалл түзеді. Бірақ мұзға айналмайды. Себебі молекулалардың ара қашықтықтары өте алыс. Сирек кездесетін жай — жаз күндерінде бұршақтың жаууы. Бұл ауаның белгілі бір қабаттарында температураның тез төмендеп, қысымның күрт көтерілуімен байланысты. Бірақ бұл құбылыс табиғатта сирек кездеседі.

ЛИТЕРАТУРА

1. Абай тілі сөздігі (словарь языка Абая). Алматы, 1968. -734 б.
2. Абрамян Л. А. Значение как категория семантики//Вопросы философии. 1965. № 1. - С. 56-66.
3. Автоматизация анализа научного текста. Киев: Наукова думка, 1984. - 258 с.
4. *Агроскина С. Н.* Структура абзаца в художественной прозе: Автореф. дис. канд., - Л.: 1972. - 25 с.
5. *Адмони В. Г.* Введение в синтаксис современного немецкого языка. - М.: Изд. лит. на иностр. языках, 1955. - 392 с.
6. *Адмони В. Г.* Синтаксическая семантика — это семантика синтаксических структур. - В кн.: Проблемы синтаксической семантики. Материалы научной конференции. - М.: Наука, 1976. - С. 3-8.
7. *Алексеев П. М.* Квантитативные аспекты речевой деятельности. - В кн.: Языковая норма и статистика. М.: Наука, 1977. - С. 43-58.
8. *Алексеев П. М., Нехай О. А.* Статистические характеристики единиц английского языка в текстах по радиоэлектронике. - В сб.: Энтропия языка и статистика речи. - Минск: 1966. С. 208-244.
9. *Алина Г. М.* Каузативные глаголы казахского и русского языков. В сб.: Проблемы сопоставительной грамматики и перевода. Алматы: АГУ, 1997. С. 23-28.
10. *Анохин П. К.* Опережающее отражение действительности.//Вопросы философии, 1962, (7. С. 14-25.
11. *Апресян Ю. Д.* Экспериментальное исследование семантики русского глагола. - М.: Наука, 1967. - 252 с.
12. *Апресян Ю. Д.* К построению языка для описания синтаксических свойств слова. - В сб.: Проблемы структурной лингвистики. М., 1973. С. 279-375.
13. *Арутюнова Н. Д.* Предложение и его смысл. - М.: Наука, 1976. - 383 с.
14. *Ахабаев Ә.* Есімдіктердің морфологиялық құрылымына статистикалық талдау. - В сб.: Статистика казахского текста. Алма-Ата: Наука, 1973. С. 88-117.
15. *Ахабаев Ә.* Қазіргі қазақ газеттері тілінің алфавитті-жиілік сөздігі. - В сб.: Статистика казахского текста. Алма-Ата: Наука, 1973. С. 344-465.
16. *Ахабаев Ә., Бектаев Ә. Б.* Зат есім сөздердің морфологиялық құрылымына статистикалық талдау. В сб.: Статистика казахского текста. Алма-Ата: Наука, 1973. С. 49-88.
17. *Ахманова О.С.* Словарь лингвистических терминов. -М.: Сов. энциклоп., 1969. - 606 с.
18. *Ахметжанова З. К.* Функционально-семантические поля русского и казахского языков. - Алма-Ата: Наука, 1989. - 106 с.
19. *Ахметжанова З.К.* Принципы сопоставительного исследования казахского и русского языков: Дис. док. - Алма-Ата, 1989. - 288 с.
20. *Ашимбетова Р.Д.* Об изучении глаголов восприятия в русском и казахском языках. - В сб.: Проблемы сопоставительной грамматики и перевода. Алматы, АГУ, 1997. С. 12-17.
21. *Әбілқасымов Б. А.* Алғашқы қазақ газеттері тілінің алфавитті-жиілік сөздігі. - В сб.: Статистика казахского текста. Алма-Ата: Наука, 1973. С. 553-566.
22. *Бадаева Н. П.* Синтаксическая структура рассказа И. С. Тургенева "Лес и степь". - В кн.: Вопросы изучения русского языка. - Ростов-на-Дону: 1964. С. 126-129.
23. *Байтұрсынов А.* Тіл тағылымы. - Алматы: Ана тілі, 1982. - 447 б.
24. *Балакаев М. Б.* Основные вопросы синтаксиса простого предложения в современном казахском литературном языке: АДД., Алма-Ата, 1950.
25. *Балакаев М. Б.* Основные типы словосочетаний в казахском языке. Алма-Ата, 1957.
26. *Балакаев М. Б.* Современный казахский язык. Синтаксис словосочетания и предложения. - Алма-Ата: АН Каз ССР, 1959. - 235 с.
27. *Барт Ролан.* Лингвистика текста. - В кн.: Новое в зарубежной лингвистике. Выпуск 8. Лингвистика текста. - М.: Прогресс, 1978. С. 442-449.
28. *Бектаев К. Б.* Статистико-информационная типология тюркского текста. Алма-Ата: Наука, 1978. - 183 с.

29. *Бектаев Қ. Б., Жұбанов А. Х., Мырзабеков С., Белботаев А. Б.* "М.О.Әуезовтің "Абай жолы" романының жиілік сөздігі". - Алматы: Ғылым, 1979. - 336 б.
30. *Бектаев К. Б.* Частотный словарь языка Абая/Межвузовская конференция по вопросам частотных словарей и автоматизации лингвистических работ. Тезисы докладов и сообщений. - ЛГУ, 1966. С. 36-37.
31. *Бектаев Қ. Б., Жұбанов А. Қ., Мырзабеков С., Белботаев А. Б.* "М.О.Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздіктері". - Алматы-Түркістан: 1995. - 346 б.
32. *Бектаев Қ. Б.* Қазақ ертегілері тілінің алфавитті-жиілік сөздігі. - В сб.: Статистика казахского текста. Алма-Ата: Наука, 1973. С. 465-519.
33. *Бектаев Қ. Б., Белботаев А. Б., Молдабеков Қ.* "F. Мүсіреповтің "Кездеспей кеткен бір бейне" повесі тілінің алфавитті-жиілік сөздігі". - В сб.: Статистика казахского текста. - Алма-Ата: Наука, 1973. С. 519-543.
34. *Бектаев Қ. Б., Белботаев А. Б.* "Өскен өркен" романының алфавитті-жиілік сөздігі. - Шымкент: 1988. - 77 б.
35. *Бектаев Қ. Б., Молдабеков Қ.* Бастауыш класс оқулықтары тексінің сөздігі. - Шымкент: 1990. - 124 б.
36. *Бектаев Қ. Б., Мырзабеков С.* Етістіктің форма өзгерткіш аффикстерінің статистикалық талдауы. - В сб.: Статистика казахского текста. - Алма-Ата: Наука, 1973. С. 136-196.
37. *Бектаев Қ. Б., Мырзабеков С.* Сөз формалары. В сб.: Статистика казахского текста. - Алматы: Наука, 1973. С. 197-224.
38. *Белботаев А. Б.* Математика тексінің алфавитті-жиілік сөздігі. - В сб.: Статистика казахского текста. Алма-Ата: Наука, 1973. С. 543-552.
39. *Белботаев А. Б.* Лингвистические характеристики частей речи казахского текста: Автореф. дис. канд. филолог. наук. - Алма-Ата, 1992. - 32 с.
40. *Белботаев А.Б., Жұбанов А.Қ., Серазитдинов З.А.* Түркі тілдеріндегі ықтималды-статистикалық әдістер негізінде жүргізілген лингвистикалық зерттеулер туралы. - В кн.: Статистика казахского текста. - Алма-Ата: Ғылым, 1990. Вып. II. С. 21-31.
41. *Белоногов Г.Г., Кузнецов Б.А.* Языковые средства автоматизированных информационных систем. - М.: Наука, 1983. - 287 с.
42. *Берзон В. Е., Зубов А. В.* О семантической классификации параметров связности текста. - В сб.: Романское и германское языкознание. Вып. I. Вопросы экспериментальной фонетики и прикладной лингвистики. - Минск: Изд. МГПИИЯ, 1977. С. 185-197.
43. *Богданов В. В.* Семантико-синтаксическая организация предложении. - Л:. ЛГУ, 1977. - 204 с.
44. *Богданов В. В.* Контекстуализация предложения. В сб .: Предложение и текст: семантика, прагматика и синтаксис. - Л.: ЛГУ, 1988. С. 23-28.
45. *Богородицкий В. А.* Общий курс русской грамматики. 5-е изд, М. - Л., 1935.
46. *Бурсак Т. Д., Шмарина В. И.* Факторы, обусловливающие информативную значимость прилагательного в тексте. - В кн: Семантические вопросы микро - и макросинтаксиса. Хабаровск, 1980. С. 27-35.
47. *Линдсей П., Норман Д.* Переработка информации у человека. - М.: Мир, 1974. - 550 с.
48. *Бергельсон М. Б., Кибрик А. Е.* Прагматический принцип приоритета и его отражение в грамматике языка. - Изв. АН СССР. сер. лит. и языка, т. 40, (4, 1981. С. 343-355.
49. *Блюменау Д. И.* Проблемы свертывания научной информации. - Л.: Наука, 1982. - 166 с.
50. *Большунов Я. В.* Воспроизведение семантически неравнозначных частей связного текста//Вопросы психологии. 1977. (ғ2, С. 114-120.
51. *Бондаренко Г. В., Шрейдер Ю. А.* Текст-смысл-ситуация (к постановке проблемы). - В кн.: Вопросы информационной теории и практики, № 36.- М.: ВИНИТИ, 1978. С. 80-91.
52. *Бондарко А. В.* Грамматическое значение и смысл. Л., 1978. - 227 с.
53. *Бондарко А. В.* Принципы функциональной грамматики и вопросы аспектологии. - Л.: Наука, 1983. - 208 с.
54. *Борель Эмиль.* Вероятность и достоверность. - М.: Наука, 1964. - 119 с.
55. *Бородаченко О. А.* Межфразовые связи в различных стилях письменного языка. - В кн.: Математическая лингвистика. Киев: КГУ, - 1973. Вып. I. С. 15-25.
56. *Брандес О. П.* Прагматика языка как переводческая проблема. - Тетрады переводчика. М., 1979. Вып. 16. С. 65-71.

57. *Важенина В. П.* К вопросу о лингвистических предпосылках смыслового анализа текста на языковых факультетах. - В кн.: Проблемы лексической и грамматической семасиологии. Владимир, 1974. С. 194-205.
58. *Васильев Л. М.* Речевой смысл предложения как результат взаимодействия языкового и неязыкового знания. - В сб.: Предложение и текст: семантика, прагматика и синтаксис. - Л., ЛГУ. 1988. С. 5-10.
59. *Васильев Л. М.* Семантика русского глагола. М.: Высшая школа, 1981. - 184 с.
60. *Венскович М. С.* Понятие ситуации и смысл высказывания. - Уч. зап. МГПИИЯ им. М. Тореза, вып. 95. Вопросы романо-германской филологии. - М., 1976. С. 27-41.
61. *Верещагин Е. М., Костомаров В. Г.* Лингвострановедческая теория слова. - М., Русский язык, 1980. -320 с.
62. *Винокур Г. О.* Избранные работы по русскому языку. - М.: Учпедгиз, 1959. - 492 с.
63. *Виноградов В. В.* Русский язык (грамматическое учение о слове). Изд. 2-е. - М.: Высшая школа, 1972. - 614 с.
64. *Вольф Е. М.* Грамматика и семантика прилагательного. - М.: Наука, 1978. - 200 с.
65. *Выготский Л. С.* Проблема сознания. - В кн.: Психология грамматики. - М.: Изд. МГУ, 1968. С. 182-195.
66. *Вятютнев М. Н.* Теория учебника русского языка как иностранного. - М., Русский язык, 1984. - 144 с.
67. *Вятютнев М. Н.* Понятие языковой компетенции в лингвистике и методике преподавания иностранных языков//Иностранные языки в школе. 1975. (№ 6, С.55-64.
68. "Қазақстан мұғалімі". Алматы, 1986. (35 статей, 500 абз.).
69. *Гальперин И. Р.* О понятии текст//Вопросы языкознания. 1974. № 6.
70. *Гальперин И. Р.* Текст как объект лингвистического исследования. - М.: Наука, 1981. - 139 с.
71. *Гаузенблаз К.* О характеристике и классификации речевых произведений. - В сб.: Новое в зарубежной лингвистике. Лингвистика текста. М.: 1978. Вып. 8. С. 57-78.
72. *Гаазе-Рапопорт М. Г.* Поиск вариантов в сочинении сказок. - В кн.: Зарипов Р. Х. Машинный поиск вариантов при моделировании творческого процесса. М.: Наука, 1983. - С. 213-223.
73. *Гак В. Г.* О семантическом инварианте предложения. - В кн.: Проблемы синтаксической семантики. Материалы научной конференции. - М.: МГПИИЯ им. М.Тореза, 1976. - С. 22-26.
74. *Гак В. Г.* Высказывание и ситуация. - В кн.: Проблемы структурной лингвистики. 1972. - М.: 1973. - С. 358-370.
75. *Гак В. Г.* К типологии лингвистических номинаций. - В кн.: Языковая номинация. - М.: Наука, 1977. - С. 230-293.
76. *Гак В. Г.* Повторная номинация на уровне предложения. - В кн.: Синтаксис текста. - М.: Наука, 1979. - С. 91-102.
77. *Гендина Н. И.* Выявление аспектов содержания документов с помощью формальных текстовых признаков. //Научно-техническая информация. Сер. 2. 1979, № 4, - С. 7-12.
78. *Гендина Н. И.* Выявление аспектов содержания документов с помощью формальных текстовых признаков. //Научно-техническая информация. Сер. 2. 1979, № 4, - С. 7-12.
79. *Герд А. С.* Логико-понятийная структура научного текста и его семантика. - В кн.: Вопросы семантики. Калининград, 1978. - С. 55-58.
80. *Гиндин С. И.* Онтологическое единство текста и виды внутритекстовой организации. - В кн.: Машинный перевод и прикладная лингвистика. - М., 1971. Вып. 4. - С. 130-145.
81. *Гриб В. В.* Проблема взаимосвязи образа и знака. - Минск: Наука и техника, 1978. - С. 166.
82. *Гриценко И. М.* Композиционно-стилистические функции абзацев в репортажах Э. Э. Киша. Автореф. дис. ... канд. филол. наук. - Л., 1966. - 22 с.
83. *Гришина О. Н.* Проблемы контекстно-вариативного членения текста в стиле языка художественной и научной прозы. - В кн.; Функциональные стили и преподавание иностранных языков. - М., Наука, 1982. - С. 52-68;
84. *Глаголев Н. В.* Семантико-структурные элементы актуализации сообщения в сверхфразовом единстве.- В кн: Проблемы синтаксической семантики. Материалы научной конференции. - М., МГПИИЯ им. М. Тореза, 1976. - С. 104-107.
85. *Головин Б. Н.* Язык и статистика. - М., 1971. - 190 с.
86. *Горбачев С.Б., Клещев А.С., Черняховская М. Ю.* Обзор языка предоставления знаний МЕДИФОР- 2. - В кн.: Теоретические основы компиляции. - Новосибирск, 1980. - С. 78-90.

87. *Горелов И. Н.* Проблема функционального базиса речи в онтогенезе. - Челябинск, 1974. - 116 с.
88. *Гулыга Е. В.* Сложно-подчиненные предложения и декоративные значения. - В кн.: Вопросы романо-германской филологии: синтаксическая семантика. - М., МГПИИЯ им. М. Тореза, 1977. - С. 51-60.
89. *Гумбольдт Вильгельм Фон.* Избранные труды по языкознанию. - М., Прогресс, 1984. - 397 с.
90. *Денисов П. Н.* Принципы отбора лексики для учебных словарей. - В кн: Вопросы учебной лексикографии. - М., Изд. МГУ, 1969. - С. 15-39.
91. *Джубанов А. Х.* Квантитативная структура казахского текста. - Алма-Ата, Наука, 1987. - 146 с.
92. *Джубанов А. Х.* Использование ЭВМ для выделения опорных слов текста. - В сб. Актуальные проблемы компьютерной лингвистики. Тезисы докладов Всесоюзной конференции в гор. Тарту 29-31 мая 1990 г. - Тарту, 1990. - С. 49-51.
93. *Джубанов А. Х., Кажибеков Е. З., Джунусбеков А. Д.* Создание словарной базы данных обще - и межтюркских моносиллабов: к проблеме разработки машинного фонда тюркских языков. // Третья Всесоюзная конференция по созданию машинного фонда русского языка. Тезисы докладов. Ч. II. - М., 1989. - С. 129-132.
94. *Джубанов А.Х., Зубов А.В.* Формализация процедуры построения текстов естественного языка. - В кн.: Научные приборы и автоматизация научных исследований. - Алма-Ата: Ғылым, 1992. - С. 190-196.
95. *Дуйсекова К. К.* Семантика глаголов ориентированного положения в пространстве в казахском языке: Автореф. дис. канд.филол. наук - Алма-Аты, 1996. - 26 с.
96. *Дүйсенбекова Л. С.* Қазіргі қазақ тіліндегі бұйрық райдың парадигмасы мен мағыналық құрамы: Канд. дис. автореф. - Алматы, 1998. - 24 б.
97. *Джанашина Н.* Экзистенциональные конструкции с Var/yjk в турецком языке. // Советская тюркология. 1987. № 6, - С. 60-77.
98. *Ельмслев Луи.* Пролегемоны к теории языка. - В сб.: Новое в лингвистике. Вып. 1. - М.: 1960. - С. 305.
99. *Ершов В. Н.* Текст и предложение. К определению статуса лингвистических объектов. - В кн.: Синтаксическая семантика и прогматика. - Калинин: Калинин. Гос. ун-т, 1982. - С. 11-22.
100. *Ершов А. П.* К методологии построения диалоговых систем: феномен деловой прозы. Препринт. 156. - Новосибирск: ВЦ СО АН СССР, 1979. - 24 с.
101. *Ермакова И. В.* Автосемантия на сверхфразовым уровне. - В кн.: Сборник научных трудов МГПИИЯ. Вып. 141. Лингвистика текста. - М., 1979. - С. 29-37.
102. *Есенов К.* Осложненные предложения в казахском языке. - Алма Ата: Ғылым, 1967.
103. *Есенов Қ.* Қазақ тіліндегі күрделенген сөйлемдер. - Алматы: Ғылым, 1974.
104. *Жиенбаев С.* Синтаксис мәселелері. - Алматы, 1941.
105. *Жинкин Н. И.* Речь как проводник информации. - М.: Наука, 1982. - 157 с.
106. *Жинкин Н. И.* Замысел речи. - В кн.: Планы и модели будущего в речи (материалы к обсуждению). - Тбилиси, 1970. - С. 13-15.
107. *Жубанов Х.* Из истории порядка слов в казахском предложении, - В кн.: Исследования по казахскому языку. - Алматы: Ғылым, 1999. - 416 б.
108. *Жубанов Е. Х.* У истоков идеи математизации лингвистических исследований. - В кн.: Статистика казахского текста. Вып. II. - Алма-Ата: Ғылым, 1990. - С. 12-21.
109. *Жұбанов А.* Проф. Құдайберген Жұбановтың буын теорясы және оның компьютерлік лингвистикадағы орны. - Мемлекеттік тіл: терминология, іс қағаздары мен бұқаралық ақпарат құралдарының тілі. Қ. Жұбановтың 100 жылдығына арналған ғылыми-тәжірибелік конф. матер. - Астана, 1999. - С. 298-305.
110. *Жұбанов А.* Құдайберген Жұбанов және қазақ мәтін мазмұнын статистикалық тәсілмен анықтау. - Құдайберген Жұбанов және қазақ тіл білімі: Республикалық ғылыми-теориялық конф. материал. - Алматы: Арыс, 1999. - С. 102-108.
111. *Жубанов А. К.* К проблеме семантической классификации знаменательных классов слов казахского языка. - //Известия МОН РК АН РК. Серия филолог., - Алматы: 2001, № 2. - С. 13-30.
112. *Жубанов А. К.* Статистическое исследование казахского текста на уровне абзацев. - // "Тілтаным - Языкознание" ИЯ им. А. МОН РК. - Алматы: 2001. № 2. - С. 35-41.
113. *Засорина Л. Н.* Частотные словари и вопросы лексико-статистики. - //Межвузовская конференция по вопросам частотных словарей и автоматизации. Л., 1966. - С. 3-4.

114. *Зарубина Н. Д.* О психолингвистическом обосновании приемлемости предложения и сверхфразового единства в качестве единиц обучения иностранному языку. - В кн.: Психология грамматики. - М., Изд. МГУ, 1968. - С. 145-152.
115. *Закиев М. З.* Актуальные проблемы в сложных предложениях в тюркских языках. - Баку, 1983.
116. *Зарубина Н. Д.* Текст: лингвистический и методический аспекты. М.: Русский язык, 1981. - 113 с.
117. *Звегинцев В. А.* Социальное и лингвистическое в социолингвистике. // Изв. АН СССР. Сер. лит. и языка. 1982, Т. 41, № 3, - С. 250-258.
118. *Звегинцев В. А.* О цельнооформленности единиц текста. // Изв. АН СССР. Сер. лит. и языка, 1980, Т. 39, № 1, - С. 13-21.
119. *Звегинцев В. А.* Язык и лингвистическая теория. - М.: Изд. МГУ, 1973. - 248 с.
120. *Звегинцев В. А.* Предложение и его отношение к языку и речи. - М.: Изд. МГУ, 1976. - 307 с.
121. *Золотова Г. А.* Очерки функционального синтаксиса русского языка. - М., 1973. - 170 с.
122. *Золотова Г. А.* О структуре простого прелложения в русском языке. // ВЯ, 1967, № 6. - С. 85-97.
123. *Зубов А. В.* Переработка текста естественного языка в системе "человек-машина". - Л.: Наука, 1971. - С. 286-434.
124. *Зубов А. В.* Программирование лингвистических задач на языке ПЛ/1. Учебное пособие. - Минск: Изд., МГПИИЯ, 1978. - 120 с.
125. *Зубов А. В.* Статистической аспект содержания текста и его формальное представление. В кн: Квантитативная лингвистика и автоматический анализ текстов. - Тарту, 1986. - С. 75-94.
126. *Зубов А. В.* Вероятностно-алгоритмическая модель порождения текста (семантико-синтаксический аспект). - Минск., 1985. - 399 с. (докторская диссертация, филол. наук).
127. *Зубов А. В.* Проблемы порождения текста. Часть I. Теория и алгоритмы. - Минск, 1987. - 77 с.
128. *Зубов А. В.* Проблемы порождения текста. Часть II. Вероятностно-алгоритмическая модель порождения текста. - Минск, 1989. - 63 с.
129. *Зубов А. В.* О вероятностно-алгоритмическом подходе к порождению текста. - В кн.: Экспериментальная фонетика и прикладная лингвистика. - Минск: Изд. МГПИИЯ, 1980. - С. 178-183.
130. *Ибадильдина З. К.* Эмотивная лексика казахского и русского языков: Автреф. дис. канд. - Алматы, 1997. - 26 с.
131. *Иванов М. Н.* Строение абзаца и его пунктуация. // Русский язык в школе. - М., 1950. № 3, - С. 16-20.
132. *Ившин В. Д.* Семантика единиц языка и речи и соотношение их с единицами мышления. В сб.: Реализация языковых единиц в тексте. Свердловск: Свердл. Гос. пед. ин-т, 1986. С. 3-8.
133. *Изенберг Х.* О предмете лингвистической теории текста. - В кн.: Новое в зарубежной лингвистике. Вып 8. Лингвистика текста. - М.: Прогресс, 1978. - С. 43-56.
134. *Исенгалиева В. А.* Употребление падежей в казахском и русском языках. - Алма-Ата: Наука, 1961. - 165 с.
135. *Исенгалиева В. А.* Тюркские глаголы с основами, заимствованными из русского языка. - Алма-Ата, Наука, 1966. - С. 55.
136. *Исхакова Х. Ф.* Исследования в области формальной морфологии тюркских языков: Автореф. канд. дисс. М., 1972. - 27 с.
137. *Канцельсон С. Д.* Типология языка и речевое мышление. - Л.: Наука, 1972. - 216 с.
138. *Каминскене Л. Г.* Членение текста на сверхфразовые единства в словесно-художественном творчестве: Автореф. канд. дисс. - М., 1972. - 27 с.
139. *Караулов Ю. Н.* Методология лингвистического исследования и машинный фонд русского языка. - В кн.: Машинный фонд русского языка: идеи и суждения. - М.: Наука, 1986. - С. 13-25.
140. *Қайшығұлова Ж. Т.* Қазақ мәтінінің күрделі фразалық тұтасым//абзац деңгейінде мүшеленуі. - Афтореф. дис. канд... филол., Алматы, 2001. - 25 б.
141. *Қалыбаева-Хасенова А. Қ.* Етістіктің лексика-грамматикалық сипаты. - Алматы: Ғылым, 1971. - 308 б.

142. *Караулов Ю. Н.* Лингвистическое конструирование и тезаурус литературного языка. - М.: Наука, 1981. - 366 с.
143. *Кенесбаев С.К., Бектаев К.Б., Пиотровский Р.Г.* О статистическом изучении тюркских текстов. - //Статистическое и информационное изучение тюркских языков. Алма-Ата, 1969. - С. 3-4.
144. *Келемен Я.* Текст и значение. - В сб.: Семиотика и художественное творчество. М.: Наука, 1977. - С. 104-124.
145. *Кестен С. А.* Структура абзаца и его роль в архитектонике и композиции художественного текста. - В кн.: Иностранные языки в вузах Узбекистана. Вып. 8. - Ташкент, 1975. - С. 256-264.
146. *Киселева Л. А.* Вопросы теории речевого воздействия. - Л.: ЛГУ, 1978. - 160 с.
147. *Киссен И. А.* Словарь наиболее употребительных слов современного узбекского литературного языка. Ташкент: Уқитивчи, - 1972. - 112 с.
148. *Көшербаев К.* Минералдар әлемінде. - Алматы, 1986. - 187 б.
149. *Кожевникова К.* Об аспектах связности в тексте как в целом. - Синтаксис и стилистика, под. ред. Г. А. Золотовой. - М., 1976.
150. *Кузьмичева И. А.* Развитие иноязычной речи в процессе коммуникативного обучения. - В сб.: Семантика, логика и интуиция в мыслительной деятельности человека. - М.: Педогогика, 1979. - С. 148-183.
151. *Кузнецова Э. М.* К вопросу о соотношении смысла и значения в сверхфразовом единстве. - В кн.: Лингвистика текста. Сб. научных трудов Московск. гос. пед. ин-та ин. яз. им. М. Тореза, Вып. 141, М., 1979. - С. 49-60.
152. *Купина Н. А.* Опыт системно-синтаксического анализа семантики связного текста. - В кн.: Семантика и структура предложения. Лексическая и синтаксическая семантика. Уфа: БГУ., 1978. - С. 137-141.
153. *Колшанский Г.В.* Контекстная семантика. - М.: Наука, 1980. - 148 с.
154. *Колшанский Г.В.* О языковом механизме порождения текста. // ВЯ. 1983. № 3. - С. 44-51.
155. *Кондаков Н.И.* Логический словарь-справочник.- М.: Наука. 1976. - 720 с.
156. *Курамысова К.Н.* Семантико-грамматические особенности глаголов речи// Актуальные проблемы общественных наук. - Алма-Ата, 1986. Ч. 2. - 206 с.
157. *Кильдибекова Г.А.* Функционально-семантическая категория каузативности в русском языке. //Межвузов. науч. сб.: Исследования по семантике. Семантика слова и словосочетания. - Уфа: БГУ, 1984. - с. 8-19.
158. *Клящев А.С.* Предоставление знаний. Методология, формализмы, организация вычислений и программная поддержка. - В кн.: Прикладная информатика. Вып.1. М.: Финансы и статистика, 1983. - С. 49-94.
159. *Коган А. Б., Чораян О.Г.* Вероятностные механизмы нервной деятельности. - Ростов: Изд. РГУ, 1980. - 175 с.
160. *Королев Э.И.* Системные требования к лингвистическому обеспечению автоматизированных информационных систем. //Научно-техническая информация. Сер. 2. 1981. № 10. - С. 1-5.
161. *Кибрик А.Е.* Предикатно-аргументные отношения в семантически эргативных языках. //Изв. АН СССР. Сер. лит. и языка. Т. 39. 1980, № 4. - С. 324-335.
162. *Котюрова М.П.* О связи предложений и абзацев в научной речи. - В кн.: Лингвостилистические исследования научной речи. - М.: Наука, 1979. - С. 176-189.
163. Қазіргі қазақ тілі (Лексика, фонетика, грамматика). Жауапты ред.: М. Балақаев, А. Ысқақов. - Алматы: Қаз.ССР ҒА баспасы, 1954. - 563 б.
164. Қазақ тілінің грамматикасы. I. Морфология. Алматы: Ғылым, 1967.- 264 б.
165. Қазақ ертегілері. - Алматы: Балауса, 2001. - 261 с.
166. Қазақ әңгімелері. - Алматы: Мектеп, 1984. - 335 б.,
[1) *Сейфуллин С.* Бандыны қуған Хамит. (102 абз.); 2) *Майлин Б.* Қанды кек. (98 абз.); 3) *Әуезов М.* Көксерек. (97 абз.); 4) *Мұқанов С.* Әсия. (47 абз.); 5) *Мүсірепов Ғ.* Жапон балладасы. (92 абз.); 6) *Ерубаев С.* Бахыт. (53 абз.)].
167. *Қалыбеков Т.* Экология және ашық кен. - Алматы, 1988.
168. *Қасабекова Қ. Қ.* Қимыл-қозғалыс етістіктердің жасалу ерекшеліктері //Қазақстан мектебі. 1985. № 4. - 27 б.
169. *Койшибаева М.С.* К вопросу о бытийных глаголах в русском и казахском языках. - В сб.: Проблемы сопоставительной грамматики и перевода. Алматы: АГУ, 1997. - С. 46-52.

170. *Құрышжанов А. Қ., Жубанов А. Қ., Белботаев А. Б.* Куманша-қазақша жиілік сөздік. - Алматы: Ғылым, 1978. - 277 с.
171. *Лапина Л.Г.* Функционирование цельных и элементарных предложений в абзаце современной немецкой научной прозы. - В кн.: Проблемы реализации системы синтаксиса. - Пермь, 1981. - С. 45-54.
172. *Ларина А.А.* О предметно-логическом содержании абзацев. - В кн.: Проблемы внутренней динамики речевых норм. - Минск: Изд. МГПИИЯ, 1982 - С. 203-209.
173. *Леонтьева Н.Н.* Семантика связного текста и единицы информационного анализа// НТИ. Сер. 2, 1981, № 1, - С. 21-29.
174. *Леонтьев А.А.* Общественные функции языка и его функциональные эквиваленты. - В кн.: Язык и общество. М., 1968.
175. *Леонтьев А.А.* Психолингвистические единицы и порождение речевого высказывания. - М.: Наука, 1969. - 307 с.
176. *Леонтьев А.А.* Высказывание как предмет лингвистики, психолингвистики и теории коммуникации. - В кн.: Синтаксис текста. отв. ред. Г.А. Золотова. М., 1979.
177. *Лесохин М.М., Лукьяненков К.Ф., Пиотровский Р.Г.* Введение в математическую лингвистику. - Минск: Наука и техника, 1982. - 263 с.
178. *Лекомцев Ю. К.* Психическая ситуация, предложение и семантический признак. - В сб.: Труды по знаковым системам, 4 уч. записки Тартус. гос. унив. Вып. 308. Тарту, 1973. - С. 5-44.
179. *Левковская Н. А.* В чем различие между СФЕ и абзацем // Филологические науки. М., 1980. № 1, - С. 32-47.
180. *Лейкина Б. М.* К проблеме взаимодействия языковых и неязыковых значений при осмыслении речи. -В кн.: Лингвистические проблемы функционального моделирования речевой деятельности. Вып. 2. ЛГУ, 1974. - С. 97-109.
181. *Лесскис Г. А.* К вопросу о грамматических различиях научной и художественной прозы (на материале русской прозы 60-х годов XIX века). - В кн.: Труды по знаковым системам. 2. Уч. зап. Тартуск. гос. унив. Вып 181. - Тарту: 1965. - С. 76-87.
182. Лингвистические вопросы алгоритмической обработки сообщений. - М.: Наука, 1983. - 246 с.
183. *Лосева Л. М.* Как строится текст. Пособие для учителей. - М.: Просвещение, 1980. - 94 с.
184. *Ляховицкая Л. М.* О некоторых параметрах связи между абзацами. - В сб.: Математическая лингвистика. Вып. I. - Киев, 1973. - С. 35-40.
185. *Маманов И. Е.* Вспомогательные глаголы в казахском языке. - Алма-Аты, 1949. - 4 с.
186. *Маманов И.Е.* Қазіргі қазақ тілі. Етістік. - Алматы, 1966.
187. *Марусенко М.А.* О формировании словника словаря статистически устойчивых научно-технических терминов. - В кн.: Структурная и прикладная лингвистика. Межвузовский сборник. Вып.2. Л.: ЛГУ, 1983. - С. 82-89.
188. Маркс К. и Энгельс Ф. Немецкая идеология. - Соч. Изд. 2. Т.3.
189. *Маслов Б. А.* Проблемы лингвистического анализа связного текста (межфразовый уровень). Учебное пособие к спецкурсу. - Таллин, 1975. - 104 с.
190. *Марчук Ю. Н.* Вычислительная лексикография. - М.: ВЦП, 1976. - 183 с.
191. *Марчук Ю. Н.* Проблемы машинного перевода. - М.: Наука, 1983. - 233 с.
192. *Маров В. Н.* Теория абзаца в советской лингвистике//Русский язык в школе. 1973. № 1, - С. 105-110.
193. *Мешман Л. И.* Композиционно-смысловая организация текста английской научной статьи. - В кн.: Функциональные стили и преподование иностранных языков. - М.: Наука, 1982. - С. 14-27.
194. *Мельничук А. С.* Взаимодействие грамматических единиц различных уровней в рамках предложения. - В кн.: Единицы разных уровней грамматического строя языка и их взаимодействие. - М., 1969.
195. *Мельников Г.П.* Системология и языковые аспекты кибернетики. - М.: Советское радио, 1978. -368 с.
196. *Мельников Г. П.* Системология и языковые аспекты кибернетики. - М.: Советское радио, 1978. - 368 с.
197. *Минский М.* Фреймы для предоставления знаний. /Пер. с англ. - М.: Энергия, 1979. - 152 с.
198. *Мильчин А.Э.* Методика редактирования текста. -М.: Книга, 1980. - 318 с.

199. *Москальская О.И.* Грамматика текста. - М.: Высшая школа, 1981. - 183 с.
200. *Мырзабеков С.* Туынды түбір етістіктердің құрылымына статистикалық талдау. - В сб.: Статистика казахского текста. - Алматы: Наука, 1973. - С. 117-135.
201. *Налимов В.В.* Вероятностная модель языка. - М., 1974. - 272 с.
202. *Невежина З.В.* Структурно-семантическая орканизация сверхфразового единства в современном английском языке: Автореф. дис. ... канд. филол.наук. - Киев, 1971. - 22 с.
203. *Нелюбин Л.Л.* Перевод и прикладная лингвистика. - М.: Высшая школа, 1983. - 207 с.
204. *Нечаева О.А.* Функционально-смысловые типы речи (описание, повествование, рассуждение). - Улан-Уде, 1974. - 262 с.
205. *Николаева В.П.* Структурно-композиционные особенности начальных и конечных абзацев. - В кн.: Лингвистика текста. Материалы научной конференции. Часть 1. - М., 1974. - С. 206-209.
206. *Никитин М.В.* О предмете и понятиях комбинаторной семантики (взаимодействие лексических значений слов в словосочетаниях). - В кн.: Проблемы лексической и грамматической семасиологии. - Владимир, 1974. - С. 3-155.
207. Новое в зарубежной лингвистике. Выпуск 8. Лингвистика текста. - М.: Прогресс, 1978. - 479 с.
208. *Новиков Э. А., Чистякова Г. Д.* К вопросу о теме и денотате//Известия АН СССР. Серия литературы и языка, том 40. М., 1981. № 1. - С. 48-56.
209. *Новиков А. И.* Семантика текста и ее формализация. - М.: Наука, 1983. - 216 с.
210. *Ноздрина Л. А.* Языковые средства делимации текста коротких рассказов. - В кн.: Лингвистика текста. Сборник научных трудов Моск. Гос. пед. ин-та ин. языков им. М.Тореза. Вып.141. - М., 1979. - С. 116-129.
211. Общее языкознание. Внутренняя структура языка. -М.: Наука, 1972. - 565 с.
212. *Оразов М.* Семантика казахского глагола (Опыт семантической классификации.): Дис. докт. - Алма-Ата, 1983. - 415 с.
213. *Оразов М.* Қазіргі қазақ тіліндегі қалып етістіктері. - Алматы, 1980.
214. *Оразов М.* Сөз теориясы. - Ташкент, 1985.
215. *Оразов М.* Қазақ тілінің семантикасы. - Алматы: Рауан, 1991. - 215 б.
216. *Оралбаева Н.* Қазақ тіліндегі етістіктің категориялары. Тіркесімі, сөздігі, орны. - Алматы, 1980. - 59 б.
217. *Оралбаева Н.* Қазіргі қазақ тілі морфологиясы. -Алматы, 1982. - 73 б.
218. *Оралбаева Н.* Қазіргі қазақ тіліндегі етістіктің аналитикалық форманттары. - Алматы: Мектеп, 1975. - 136 б.
219. *Отрадинская Л. Б.* О некоторых семантико-синтаксических особенностях организации текста научной прозы. - В кн.: Иностранные языки в вузах Узбекистана. Сб. научных трудов, (587. Ташкент: 1979. - С. 108-115.
220. *Орфеев Ю. В., Тюхтин В. С.* Мышление человека и "искусственный интеллект". - М.: Мысль, 1978. - 149 с.
221. *Османов М.-Н.* О. Частотный словарь Унсури. - М., 1970. - 320 с.
222. Основы построения описательной грамматики современного русского языка/Под ред. Н. Ю. Шведовой. - М., 1966. - 270 с.
223. *Падучева Е. В.* О структуре абзаца. - В кн: Труды по знаковым системам. 2. Учен. зап. Тартуск. гос. унив. Вып. 181. Тарту, 1965. - С. 284-292.
224. *Панов Е. Н.* Знаки, символы, языки. - М.: Знание, 1980. - 192 с.
225. *Перекальская Т. К.* Абзац как основная единица научного текста. - В кн.: Вопросы методики и лингвистики М.: Изд. МГУ, 1980. - С. 8-10.
226. *Пешковский А. М.* Русский синтаксис в научном освещении. Изд. 7- ое. - М.: Учпедгиз, 1956. - 511 с.
227. *Перебейнос В. И.* Некоторые проблемы порождения текста. - В кн.: Всесоюзный симпозиум по логике и методологии науки. Киев: октябрь 1976 г. Тезисы сообщений. - Киев: Наукова думка, 1976. - С. 170-171.
228. *Пиотровский Р. Г.* Текст, машина, человек. - Л.: Наука, 1975. - 327 с.
229. *Пиотровский Р.Г.* Инженерная лингвистика и теория языка.- Л.: Наука, 1979. - 112 с.
230. *Пиотровский Р. Г.* Информационные измерения печатного текста. - В сб.: Энтропия языка и статистика речи. - Минск, 1966 - С. 5-86.
231. *Пиотровский Р. Г.* Отраслевой вероятностный машинный перевод. - В сб.: Статистика текста. Том 2. - Минск, 1970. - С. 5-32.
232. *Плотников Б. А.* Основы семасиологии. - Минск: Вышэйшая школа, 1984. - 223 с.

233. Прогноз в речевой деятельности. - М.: Наука, 1974. - 213 с.
234. *Попов Э. В.* Общение с ЭВМ на естественном языке. - М.: Наука, 1982. - 360 с.
235. *Покусаенко В. К.* К вопросу о функциях абзаца. - В кн.: Структура предложения и абзац. Ростов-на-Дону, 1974. - С. 112-121.
236. *Поспелов Н. С.* Сложное синтаксическое целое и основные особенности его структуры. - Доклады и сообщения института русского языка АН СССР. 1948. Вып. 2. - С. 43-69.
237. Проблемы текстуальной лингвистики. - Киев: Изд. КГУ, 1983. - 175 с.
238. *Пропп В.* Морфология сказки. - М.: Наука, 1969. - 168 с.
239. *Разинкина Н. М.* О понятии стереотипа в языке научной литературы. - В кн.: Проблемы повышение эффективности и качество обучения иностранным языком научных работников/ Тезисы докладов. Москва-Цхалтубе, 1978. - С. 32-34.
240. *Реферовская Е.А.* Коммуникативная структура текста. - Л.: Наука, 1989. - 168 с.
241. *Реферовская Е.А.* Лингвистические исследования структуры текста. - Л., 1983. - 207 с.
242. *Рошияну Николае.* Традиционные формулы сказки. - М.: Наука, 1974. - 216 с.
243. *Розов Н.А.* Информационно-семиотические исследования: процессы-эстафеты и принцип дополнительности//-Научно-техническая информация.1984. № 2. - С. 1-7.
244. *Румянцева Э.А.* Инженерно-математический стиль мышления в современной науке. - Минск: Вышэйшая школа, 1978. - 150 с.
245. Русская грамматика. Том I. - М.: Наука, 1980. - 783 с.
246. Русская грамматика. Том II. - М.: Наука, 1980. - 709 с.
247. *Рыбников Н. А.* Язык ребенка. 2-е изд. М.- Л.: 1926.
248. *Рылова Т. Н.* Принципы упорядочения и группировка значимых слов текста для совершенствования критерия семантической связи между предложениями// Научно-техническая информация. 1973. № 4. - С. 34-38.
249. *Сауранбаев Н. Т.* Қазақ тіліндегі құрмалас сөйлемдер жүйесі. - Алматы, 1948.
250. *Сауранбаев Н. Т.* Семантика и функции деепричастия в казахском языке. - Алма-Ата, 1944. Отв. ред. Жиенбаев С.
251. *Салмина Н. Г.* Знак и символ в обучении. - М.: МГУ, 1988. - 288 с.
252. *Седов Е. А.* Эволюция и информация. - М.: Наука, 1976. - 232 с.
253. Семиотика. - М.: Радуга, 1983. - 636 с.
254. *Серкова Н. И.* Семиотика абзаца в научной и художественной литературе. - В кн.: Семантические вопросы микро- и макро-синтаксиса. - Хабаровск, 1980. - С. 27-35.
255. *Сильман Т. И.* Структура абзаца и принципы его развертывания в художественном тексте. - В кн.: Теоретические проблемы синтаксиса современных индо-европейских языков. - Л.: Наука, 1975. - С. 208-216.
256. *Сильман Т. И.* Проблемы синтаксической стилистики.- Л., 1967.- 230 с.
257. *Сильдмяэ И.* Гносеологическая структура текста и знаний. - В кн.: Семантика и представлений знаний. Труды по искусственному интеллекту. Часть 2. - Тарту, 1980. - С. 108-117.
258. *Слюсарева Н. А.* Теория Ф. де Соссюра в свете современной лингвистики. - М.: Наука, 1975. - 112 с.
259. *Соссюр Ф. де.* Курс общей лингвистики. - М., 1933.
260. *Соссюр Ф. де.* Труды по языкознанию. - М.: Прогресс, 1977. - 696 с.
261. *Солнцев В. М.* Язык как системно-структурное образование. - М.: Наука, 1977. - 342 с.
262. *Смирницкий А. И.* Лексикология английского языка - М., 1956.
263. *Супрун А. Е.* Лекции по языковедению. - Минск: БГУ, 1978. - 144 с.
264. *Степанов Г. В.* Цельность художественного образа и лингвистическое единство текста. - В кн.: Лингвистика текста. Материалы научной конференций. Часть 2. - М., 1974. - С. 72-76.
265. *Степанов Ю. С.* Семантика. - М., 1971. - 168 с.
266. *Степанов Ю. С.* В поисках прагматики (проблемы субъекта)//Изв. АН СССР. Сер. лит. и яз. Т. 40. 1981. № 4. - 325 с.
267. *Сусов И. П.* Семантическая структура предложения (на материале простого предложения современного немецкого языка). - Тула, 1973. - 142 с.
268. *Свотина М. Г.* Абзац как единица речевой практики. - В кн.: Теоретические проблемы синтаксиса современных индо-европейских языков. - Л.: Наука, 1975. - с. 205-208.
269. *Страхова В. С.* Внешние средства организации текста. - В кн.: Лингвистика текста. Сборник научных трудов МГПИИЯ им. М. Тореза. Вып. 141. М., 1971. - С. 15-159.

270. *Серкова Н.И.* Семантика и метасемантика красной строки в художествен-ной прозе. - В кн.: Актуальные вопросы лингвистического и методологического изучения художественного текста. - Владимир, 1981. - С. 108-120.
271. *Скибо В. Н.* Некоторые особенности стилистико-синтаксической организации текста (контактность абзацев). - В кн.: Лингвистика текста. Материалы научной конференции. Часть 2. М., 1974. - С. 57-59.
272. *Солганик Г.Я.* Синтаксическая стилистика.- М.: Высшая школа, 1973. - 214 с.
273. *Смирнов А. А.* Проблему психологи памяти. - М.: Просвещение, Наука, 1966. - 423 с.
274. *Соловьян В. А.* Единство экстра- и интралингвистического при интеграции текста. - В кн: Сборник научных трудов МГПИИЯ им. М. Тореза. Вып. 141. Лингвистика текста. - М., 1979. - С. 138-150.
275. *Сулейменова Э. Д.* Семантические параметры слова в интерлингвальном исследовании (теоретические и прикладные аспекты контрастивного анализа): АДД., М., 1992.
276. Сопоставительная грамматика русского и казахского языков. Морфология. - Алма-Ата: Наука, 1966. - 458 с.
277. Современный русский язык. Морфология (Курс лекций). - М., 1952. - 233 с.
278. Современный казахский язык. Алма-Ата: - Изд. АН КазССР. 1962.- 453 с.
279. *Степанов Ю. С.* Основы общего языкознания. - М.: Просвещение, 1975. - 271 с.
280. *Скороходько Э. Ф.* Семантические сети и автоматическая обработка текста. - Киев: Наукова думка, 1983. - 218 с.
281. *Смит ДЖ. Б.* Тематическая структура и тематическая сложность. - В кн.: Новое в зарубежной лингвистике. Вып. IX. - М.: Прогресс, 1980. - С. 333-355.
282. Теория и практика английской научной речи/Под редакцией М. М. Глушенко. - М.: МГУ, 1987. - 240 с.
283. *Тенишев Э. Р.* Глаголы движения в тюркских языках. - В кн.: Исторические развитие лексики тюркских языков. - М.: АН СССР, 1961. - С. 232-293.
284. *Токтарова Т.Ж.* Семантика глаголов созидания казахского и русского языков: Автореф. дис. канд. - Алма-Аты, 1998. - 25 с.
285. *Токтарова Т.* Некоторые аспекты исследования семантики глаголов созидания. - В сб.: Проблемы сопоставительной грамматики и перевода. - Алматы, АГУ, 1997. - С. 28-34.
286. *Троянская Е.С.* Лингво-стилистическое исследование немецкой научной литературы. - М.: Наука, 1982. - 312 с.
287. *Турмачева Н.А.* К вопросу о природе сложного синтаксического целого и абзаца / Синтагматика, парадигматика и их взаимоотношение / Отв. ред. Ю. Каклинь. - Рига, 1970. - 300 с.
288. *Уразымбетова А.Б.* Глоголы физического восприятия русского и казахского языков и некоторые аспекты их перевода: Автореф. дис. канд.,- Алматы, 1998. - 25 с.
289. *Уфимцева А. А.* Типы словесных знаков. - М.: Наука, 1974. - 206 с.
290. *Ұйықбаев И. Қ.* Есімдіктің түрлері. Қазақ тілі методикасының мәселелері. Алматы. 1969. - 180 б.
291. *Филлмор Ч.* Дело о падеже открывается вновь. - В кн. Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика.- М.: Прогресс, 1981. - С. 496-530.
292. *Филлмор Ч.* Основные проблемы лексической семантики. - В кн.: Новое в зарубежной лингвистике. Вып. XII. Прикладная лингвистика. - М.: Радуга, - С. 74-122.
293. *Фейгенберг И. М., Иванников В. А.* Вероятностное прогнозирование и преднастройка к движениям. - М.: Изд. МГУ, 1978. - 112 с.
294. *Фридман Л. Г.* К вопросу о сверхфразовых единицах (на материале немецкого языка). - В кн.: Теоретические проблемы синтаксиса современных индо-европейских языков. - Л.: Наука, 1975. - С. 216-221.
295. Фрумкина Р. М., Штейнфельдт Э. А. Статистические методы отбора лексики для словаря-минимума по русскому языку//Русский язык в национальный школе. - М., 1960. № 6. - С. 17-25.
296. *Фрумкина Р М.* Статистические методы изучения лексики. - М., 1964. - 115 с.
297. *Фрумкина Р. М.* Вероятность элементов текста и речевое поведение. - М., Наука, 1971. - 168 с.
298. *Хабаров И. А.* Философские проблемы семиотики. -М.: Высшая школа, 1978. - 160 с.

299. *Хлебникова И. Б., Потапова Н. Ф.* Проблема лексико-семантических связей в сверхфразовом единстве. - В сб.: Реализация языковых единиц в тексте. Свердл. Гос. пед. инст., 1986. - С. 44-48.
300. *Хомский Н.* Аспекты теории синтаксиса. - М.: Изд. МГУ, 1972. - 259 с.
301. *Храковский В. С.* Проблемы деривационной синтаксической теории: Автореф. докт. дисс. - Тбилиси, 1972. - 30 с.
302. *Чахоян Л.П.* Речемыслительные процессы и значение высказывания. - В сб.: Предложение и текст: семантика, прагматика и синтаксис. - Л.: ЛГУ, 1988. - С. 10-15.
303. *Чан Нгок Тхем.* Абзац как грамматико-семантическая сущность. - В сб.: Предложение и текст: семантика, прагматика и синтаксис. - Л.: ЛГУ, 1988. - С. 118-123.
304. Частотный словарь русского языка /Под редакцией Л. Н. Зосориной/. - М.: Русский язык, 1977. - 935 с.
305. *Чесноков П.В.* Об отношении между речевым и мыслительным процессами с точки зрения единства языка и мышления. - В кн.: Синтаксическая семантика и прагматика. - Калинин: КГУ, 1982. - С. 38-47.
306. *Чернухина И. Я.* К вопросу о единицах связного текста. - В кн.: Лингвистика текста. Материалы научной конференции. Часть 2, - М., 1974. - С. 155-158.
307. *Чистякова Г.Д.* Смысловая структура текста как определяющий фактор его понимания. - В кн.: Семантика, логика и интуиция в мыслительной деятельности человека. М.: Педагогика, 1979. - С. 101-126.
308. *Чиквашивили К. С.* Структура семантической организации текста: Автореф. дис. ... канд. филол. наук. - М., 1980. -24 с.
309. *Шабес В. Я.* Событие и текст. - М.: Высшая школа, 1989. - 175 с.
310. *Шайкенова Ж. К.* Семантика глаголов движения в казахском и русском языках: Автореф. дис... канд. филол. наук. - Алматы, 1981. - 23 с.
311. *Шахнарович А. М.* Место семантического компонента в языковой способности. - В кн.: Семантика языковых единиц и текста (лингвистические и психологические исследования). - М.: 1979. - С. 5-18.
312. *Шеннон К. Е.* Математическая теория связи. - В кн.: Работы по теории информации и кибернетики. - М.: Изд. иностр. литер., 1963. - С. 243-332.
313. *Шерковин Ю.А.* Психолингвистические проблемы массовых информационных процессов. - М.: Мысль, 1973. - 213 с.
314. *Ширяев А. Ф.* Перевод как объект комплексного научного изучения. - В кн.: Лингвистические проблемы перевода. - М.: Изд. МГУ, 1981. - С. 68-79.
315. *Шингарева Е. А.* О двух направлениях представления семантики текста// Научно-техническая информация. Сер.2. 1982. № 8. С. 1-8.
316. *Шрейдер Ю. А.* Логика знаковых систем (элементы семиотики). - М., 1974. - 123 с.
317. *Шведова Н.Ю.* О значении единиц разных уровней языка//Всесоюзная научная конференция по теоретическим вопросам языкознания (тезисы докладов и сообщений пленарных заседаний). - М., 1974. - 275 с.
318. *Шрамм А.Н.* Очерки по семантике качественных прилагательных (на материале современного русского языка). - Л.: ЛГУ, 1979. - 134 с.
319. *Школьник Л.С.* Экспериментальное изучение логико- мотивационной структуры прагматически ориентированного текста. - В кн.: Тезисы 6 Всесоюзного симпозиума по психолингвистике и теории коммуникации. Москва, 17-20 апреля 1978 г. - М.: Наука, 1978. - С. 209-211.
320. *Шмидт Зигфрид.* "История" как базовые категории. - В кн.: Новое в зарубежной лингвистике. Вып. 8. Лингвистика текста. - М.: Прогресс, 1978. - С. 89-108.
321. *Шмелев Д. Н.* Синтаксическая членимость высказывания в современном русском языке. - М.: Наука, 1976. - 150 с.
322. *Шрейдер Ю. А.* Текст, автор, семантика и информатика. Вып. 7. - М.: ВИНИТИ, 1976. - С. 153-169.
323. *Щерба Л. В.* О трояком аспекте языковых явлений и об эксперименте в языкознаний. - В кн.: Звегинцев В. А. История языкознания XIX-XX вв. в очерках и извлечениях. Ч. 2. - М., 1965. - С. 361-373.
324. *Щерба Л. В.* Преподавание иностранных языков в средней школе. М. - Л., 1947.
325. *Щербак А. М.* Очерки по сравнительной морфологий тюркских языков: (глагол). - Л.: Наука, 1981. - 183 с.

326. *Щерба Л. В.* Языковая система и речевая деятельность. - Л.: Наука, 1974. - 428 с.
327. *Ысқақов А. Ы.* Қазіргі қазақ тілі. Морфология. - Алматы: Ана тілі, 1991. - 382 б.
328. *Юрченко В. С.* Значение предложений и их компонентов. - В кн.: Синтаксическая семантика и прогматика. - Калинин: КГУ, 1982. - С. 48-55.
329. *Agricola E.* Vom Text zum Thema. - Studia Grammatica XI. Probleme der Textgrammatik. - Berlin: Akademie Verlag, 1976. S.13 -27.
330. *Barthes R.* La linguistique du discours // Sign. Language. Culture. La Hague; Paris, 1970.
331. *Bastian S.* Die Rolle der Prainformation bei der Analyse und (bersetzungvon Texten. - In: (bersetzngswissenschaftliche Beitr(ge 2: Sprachliches und Aubersprachliches in der kommunikation. Leipzig: VEB Verlag Enzyklopedie, 1979. C. 90-133.
332. Building Better English. N. Y., 1968. V. 11.
333. *Cartellitri C.* Zur Analyse des Ansgangstxtes beim (bersetzen. - In: (bersetzung-swissenschaftliche Beitr(ge 2: Sprachliches und Aubersprachliches in der Kommunikation. Leipzig: VEB Verlag Enzyklopedie, 1979. - S. 9 - 45.
334. *Dijk T. Van.* Textwissenschaft. Eine interdisziplinare Einfuhrung. Tubingen, 1980.
335. *Hartmann P.* Zur antropolohischen Fundierung der Sprache. - In: Symbolae Linguisticae in honorem G. Kurlowicz. Warszawa, 1965. S. 110-119.
336. *Hartmann P.* Zum Begriff des sprachlikhen Zeichens. -Zeitschrift f(r Phonetik, Sprachwissenschaft und Kommunikations-forschunh, 1968, 21. S. 205-222.
337. *Harweg R.* Pronomina und Textkonstitution. - M(nchen, 1968. - 382 s.
338. *Hutchins W. J.* The generation of Syntactic structures from a semantic base. - London, 1971. - 197 p.
339. *Jager G.* Translation und Translations linguistic. - Haale (Saale): VEB Max Niemeyer Verlag, 1975. - 214 s.
340. *Kade O.* Die Sprachmittelung als gesellschafliche Erscheinung und Gegestand wissenschaftlicher Untersuchund. - In: (bersetzung wissenschaftlicher Beitr(ge Z. Leipzig: VEB verlag Enzyklopedie, 1980. S. 1-285.
341. *Koch W.* Einige Probleme der Textanalyse. - Lingua, 1966, vol. 16. S. 383-398.
342. *Koen F., Becker A.,* Youngr. Psychological Reality of the Paragraph. - In: Studies in Language and Language Behaviour. The Universiy of Michigan, 1968. VI, p. 117-126.
343. *Lakoff G.* On Generative Semantics. - In: Semantics. An Interdisciplinary Reader in Philosophy, Linguisciplinary Reader in Philosophy, Linguistics and Psychology. Cambridge University Press, 1971, P. 232-296.
344. *Morris Ch.* Signs, Languague and Behavior. N. Y., 1955. P. 353
345. *Mukarovsky J.* Estetika jazyka. - In: Mukarovsky J. Kapitoly ((ske poetiky. I, 1948, s. 44-86.
346. *Ogden C. K., Richards I. A.* The Meaning of Meaning. London, 1946. 363 P.
347. *Pfutze M.* Grundgedanken au einer functionalen Textlinguistik // Textlinguistik. Dresden, 1970. I.
348. *Piotrowski R.* The Linguistcs of the Text and Machine Translation. - American Journal of Computational Linguistics, 1975 (3, P. 55-56.
349. *Riesel E., Schendels A.* Deutsche Stilistik. - Moskau: Verlag Hochschule, 1975. - S. 264.
350. Russian Literature. - The Hague - Paris, 1972. № 2.
351. Semiotyka i Struktura tekstu. - Warszawa, 1973. 352. Serkova N. I. Functional Perspective as a Logical Category// Optimization of the Linguistic Message. M., 1974.
352. *Serkova N. I.* Functional Perspective as a Logical Category// Optimization of the Linguistic Message. M., 1974.
353. *Sgall P.* Meaning of Sign, Cognitive Content, and Pragmatics. // PBML, № 25, 1976.
354. *Stern W. G.* Die Kindersprache. - Leipzig, 1928. - 327 s.
355. *Tanner W. M., Chewer W. E.* English for Every Use. - Ginn Company, 1947. P. 241-255.

ҚОЛДАНБАЛЫ ЛИНГВИСТИКА: **ФОРМАЛДЫ МОДЕЛЬДЕР**



Автордың 2007 жылы жарық көрген бұл оқу құралында қолданбалы тіл білімінің жаңа саласы "Мәтін лингвистикасына" сәйкес мәтін бірліктерінің таңбалық сипаты, абзацтардың мазмұнға қатысты ерекшеліктері мен рөлі анықталады.

Қазіргі тіл зерттеу тәжірибесіндегі модельдеу әдісінің маңыздылығымен байланысты "табиғи тіл — әлем моделі" қатынасы тілдің синтаксистік және семантикалық, статистикалық модельдерін құрастыру арқылы дәйектеледі.

Қолданбалы лингвистиканың аса маңызды мәселелері ретінде — фонетикалық аспектілер мен терминтанудың негізгі бағыттары сөз болады.

• АЛҒЫ СӨЗ

• КІРІСПЕ

• *Бірінші тарау*
МӘТІН – ҚАЗІРГІ ЛИНГВИСТИКАНЫҢ НЕГІЗГІ ЗЕРТТЕУ НЫСАНЫ

• *Екінші тарау*
ТІЛДІ МОДЕЛЬДЕУ

• *Үшінші тарау*
ТАБИҒИ ТІЛДІҢ СИНТАКСИСІ МЕН СЕМАНТИКАСЫН МОДЕЛЬДЕУ

• *Төртінші тарау*
ҚОЛДАНБАЛЫ ЛИНГВИСТИКАНЫҢ ФОНЕТИКАЛЫҚ АСПЕКТІЛЕРІ

• *Бесінші тарау*
ТІЛ БІЛІМІНДЕГІ СТАТИСТИКАЛЫҚ МОДЕЛЬДЕУ ӘДІСТЕРІ

• *Алтыншы тарау*
ҚОЛДАНБАЛЫ ТЕРМИНТАНУ

• *Жетінші тарау*
ҚАЗАҚ МӘТІНІНІҢ АБЗАЦТЫҚ ҚҰРЫЛЫМЫНЫҢ ЫҚТИМАЛДЫ-СТАТИСТИКАЛЫҚ МОДЕЛІ

• ҚОРЫТЫНДЫ

• ӘДЕБИЕТ

Алғы сөз

Математикалық ойлау әдістерінің лингвистика саласына ене бастауынан-ақ, логика ғылымы алдыңғы қатарға шыға бастады және осының нәтижесінде логика-математикалық символдар жүйесі пайда болды. Бұл жүйені математикалық мазмұндағы жасанды тіл тұрғысынан ұғынған жөн. Осындай түсінік, негізінен, ғалым Г.Фрегенің атымен тығыз байланысты екендігі де мәлім.

Осылайша қазіргі формалды логика пәні тіл ғылымына қарай бетбұрыс жасады. Сол сияқты, тіл зерттеу аясында математикалық әдістің қолданылу нәтижесінде қазіргі кездегі "машиналық аударма" деп аталатын ғылым саласы өмірге келді. Осыдан кейін табиғи тілдің әртүрлі математикалық модельдерін құруға талпыныс басталды деуге болады. Мұндай әрекеттің нәтижесінде құрылымдық лингвистикаға ұқсас, бірақ өзінше бір бөлек тіл іспеттес – жасанды таңбалар жүйесі құрылды. Тіл білімі мен логика салаларының бір-бірімен жақындасу процесінің негізінде Н.Хомскийдің "туындату грамматикасы" (порождающая грамматика) пайда болды.

Мұндай екі ғылым саласын, яғни лингвистика мен логиканың өзара "жақындасу" жағдаяты тілдік зерттеулердің барлық деңгейлерінде бірдей көрініс тапты.

XX ғасырда аталған жайт семиотика саласында да кең түрде орын алып, таңбалар теориясына жалпылаушы сипат әкелді. Тілдік жүйені зерттеуде қолданылып жүрген "таза лингвистикалық" пен "таза логикалық" әдістер үшін мұндай жалпылаушы қасиет бұрын тек зерттеушілердің "арманы" ғана болатын.

Семиотикалық тұрғыдан лингвистика мен логика салалары да таңбалар жүйесін зерттейтін ғылымға айналды. Сондықтан тілдің "таңбалар жүйесін" тілдің синтаксистік, семантикалық және прагматикалық деңгейлерінде зерттеген дұрыс деп саналды. Яғни, семиотикалық көзқараспен қарастырғанда, лингвистика мен логика ғылым салаларын мазмұнның берілу тәсілін анықтайтын ғылым саласы деуге де болады. Мазмұнның берілуі табиғи сөйлеу тіліне, жасанды сөйлеу тіліне (эсперанто), ғылыми символика жүйесіне және т.б. қатысты болып келеді. Зерттеу әдістерін лингвистикалық немесе логикалық деңгейде іске асыру прагматикалық деңгейдегі семиотикаға да қатысты. Тілдің қарым-қатынас құралы болуы және оның тіл иесіне қатынасы болуы – барлық тілдерге ортақ сипат. Ал мұндай прагматикалық деңгейдегі зерттеулерде синтаксистік пен семантикалық талдау нәтижелерінің толығымен ескерілетіні белгілі.

Тілді семиотикалық тұрғыдан зерттеудің ең басты проблемасы – тілдік айтылымдардың мән-мағыналық жағы мен осы тілді қолданушының ол жайлы түсінігін анықтау. Соңғы кезде логика-семантикалық деңгейдегі "мағына мәселесіне" мынадай анықтама беріліп жүр: "Табиғи тіл мен жасанды тілдің мазмұнын (мәнін, мағынасын) анықтау дегеніміз еркін түрде алынған екі тілдік берілістің бірдей мазмұнға ие болу жағдаяты, немесе кибернетика тілімен айтқанда, олардың бірдей ақпаратқа ие болу мүмкіндігі". Әрине, бұл түсінік, шартты түрде алғанда, тек формалды-логикалық сипатты ғана қамтиды.

Семиотика-семантикалық тұрғыда алғанда семантиканың негізгі мәселесінің тұжырымдалуы мынандай сауалдарға тіреледі: таңба өзі белгілейтін "нәрсемен" (зат, құбылыс) қандай байланыста болады; таңбалар жүйесі "ой қызметінде" (интеллектуальная деятельность) қандай орын алады; қалай, неге және қандай дәрежеде бір адам екіншісінің айтқанын түсінеді; мәдениеттің даму процесінде неге жаңа және аса күрделі таңбалар жүйесі пайда болады; семантикалық тұрғыда таңбалық жүйелер қалайша бір-бірінен айырым табады және олардың биологиялық кодтан қандай айырмашылығы бар.

Әрине, ұсынып отырған оқу құралында мұндай сауалдардың барлығына бірдей жауап табу мүмкін бола бермейді. Тілдік таңбаға қатысты әрқилы семантикалық сипаттаулардың (логикалық, прагматикалық, лингвистикалық) ішінен айырықша назар аударуға тұратыны акад. Р.Г.Пиотровскийдің таңбалар жүйесіне байланысты берген инженерлік-лингвистикалық түсініктемесі. Міне, автор өз зерттеуінде осы түсініктемелерді ұстаным етіп, мәтін ішінде кездесетін тілдік бірліктердің таңбалық сипатын толығымен қамтыған.

Тіл қызметі, ең алдымен, адамдар арасындағы "тілдесу" (общение) арқылы, яғни басқа сөзбен айтқанда, "мағына" алмасуы негізінде іске асатыны белгілі. Сондықтан да қолданбалы тіл білімі саласында мағына ең маңызды мәселелердің бірі болып саналып жүр.

Компьютер арқылы іске асатын аударма проблемасындағы басты кедергі де осындай тілдік бірліктердің мағынасы мен мәтін мазмұнын дұрыс анықтау проблемасына тіреледі. Бұл тұрғыдағы басты қиындық – сөздердің мағынасын дұрыс ажырата білу мәселесі. Әсіресе, олар көпмағыналы сөздер мен омоним сөздерді және синоним сөздерді ажырата білу мәселесі. Ал екінші мәселе – мәтін ішінде орын алып мағыналық тұтастықты сақтайтын сөз тіркестерді еркін тіркестен ажырата білу. Бұл айтылғандар синтагматика-синтаксистік бөлшектеуге қатысты мәселе екендігі айқын.

Сөздер мәтін ішінде өзара қатар тұрып та немесе белгілі қашықтықта тұрып та бір мағынаға ие бола береді. Міне, осындай мағыналық бірліктерді мәтін ішінен бөліп алу түпнұсқа тілінің синтаксистік, семантикалық модельдерін құра білу ісінде аса маңызды.

Осы және басқа да қолданбалы тіл білімінің формалды-мағыналық модельдеу әдісіне қатысты мәселелер белгілі дәрежеде ұсынылып отырған осы оқу құралында көрініс тапқан.

проф. Э.Д. СҰЛЕЙМЕНОВА

КІРІСПЕ

Қазіргі қоғамдағы ақпарат ағымының үнемі ұлғайып тұру сипаты тұтынушының мүмкіндігімен өне бойы сай келе бермейді. Мәселен, Кеңес Одағы кезіндегі деректерге қарағанда, халық шаруашылығы бойынша ғана алғанда, ауыспалы ақпарат көлемінің әрбіреуі 500 бет болатын 25 миллион кітап санымен есептелінсе, жазба құжаттардың жылдық саны – 60 миллиардпен және әлемдік шеңберде, тек патенттік құжаттардың жылдық өсу қарқыны ғана 0,5 пен 1 миллион аралығында болған екен. Басқаша айтқанда, тек ғылыми-техникалық ақпарат бойынша бір ғана маман 1 минутта 2 мың баспа табақ материалмен танысуы қажет екен. Сол сияқты, әрбір педагог, инженер, ғалым өз саласы бойынша күн сайын әлемдік әдебиеттерді бақылап отыруы үшін кемінде 1500 бет көлемдегі мәтінмен танысып отыруы қажет болған [Новиков, 1974].

Әрине, мұндай көлемдегі ақпараттарды игеру адам мүмкіншілігімен сай емес екені айқын. Мысалы, Кеңес Одағы кеңістігінде барлық жаңалықтар мен жаңадан ойлап шығару ұсыныстарының тек төрттен бірі ғана соны жаңалыққа саналып, ал басқалары, негізінен, қайталанулар болып табылған [Алейник, 1974, 50-57-бб.]. Сондықтан кейбір құнды-құнды ақпараттардың мамандар назарынан тыс қалуы мемлекет деңгейіндегі экономикалық шығынға әкелетіні белгілі.

Аталған қиыншылықтармен күресу жолдарының алғашқылары ретінде елімізде ашылып жатқан кітапханалар мен ақпараттық зерттеу институттарын (ВИНИТИ, ЦНИИПИ және т.б.) атауға болар еді. Сол сияқты шет тілінен аударма жасау ісі, реферат пен аннотациялау, қажетті мәліметтерді тауып алу жұмыстары, көбінде, дәстүрлі әдістермен, яғни автоматсыз жолдармен іске асып жатқаны мәлім. Ақпараттық қызметтің дұрыс ұйымдастырылмауынан зерттеуші-ғалым уақытының 80 пайызы оған қажетті материалдарды тек іздестіру, талдау және сұрыптау үшін ғана кетеді екен. Міне, осы себепті ақпараттық қызметті жаңа жолға қою үшін табиғи тілдегі ақпараттарды өңдеудің жаңа әдістерін іздестіру заман талабы деуге болады.

Табиғи тілдегі ақпарат ағымын "тізгіндеудің" ең бір мүмкін жолы – жаңа технологияны, яғни электронды-есептеу техникасын (компьютерлерді) молынан пайдалану.

Қазіргі кездегі компьютерлер мынадай үш түрлі ақпарат жүйелері үшін таптырмайтын техникалық база бола алады:

1. Автоматтандырылған аударма жүйесі (ААЖ) және аннотациялау мен рефераттау жүйелері (АРЖ).
2. Автоматтандырылған аударма жүйесі (ААЖ) және ақпараттық-іздеу (АІЖ) жүйесі.
3. Эксперттік жүйе (ЭЖ).

Мұндай жүйелердің ұтымдылығы, негізінен алғанда, сол табиғи тілдің функционалды моделінің құрылымдық сапасына тікелей байланысты болатыны анықталып отыр [Марчук, 1983]. Әзірше табиғи тіл қызметін толық өз міндетіне алатын автоматтандырылған модель жүйесі жоқ деуге де болады.

Жоғарыда аталған үш түрлі жүйеде мәтінге қатысты орындалатын амалдар, яғни аударма жасау, рефераттау мен аннотациялау, ақпаратты іздеу мен мәтінді туындату проблемасы әр сөздің мағынасы арқылы сөйлем мән-мағынасын ашуға, ал содан кейін барып, жинақталған белгілі нәтижелер негізінде мәтін мазмұнын айқындау принциптеріне негізделеді. Бірақ сөйлемнің мағынасы оның құрамындағы сөздер мағыналарының, ал мәтін мазмұны – сөйлемдер мән-мағыналарының бірден-бір жиынтығы емес. Сондықтан аталған жүйені құруда адам қызметінің мәтінді туындатуға қатысты шығармашылық жағын модельдеуге қатысты жаңаша бір бағыт-бағдарды ұстаным ету қажет. Әрине, тіл табиғатының аса күрделік сипаты мұндай бағытты табуға кедергі келтіріп, қиындататыны белгілі. Егер табиғи тіл қызметінің формалды моделін құру синтетикалық процесс екендігін ескеретін болсақ, аталған модельді құру ісіне тілші мамандармен бірге басқа да, яғни психолингвист, психолог, логик, философ, кибернетик, математик және т.б. салалардың мамандары өзара түсінісе қызмет етуі қажет.

Тіл қызметінің формалды моделін құруға қатысты зерттеулерде мынадай негізгі үш жағдайды атап кетуге болады:

1) тілді зерттеу мен сипаттауда адресант (айтушы) пен адресат (тыңдаушы, қабылдаушы) факторы есепке алынбайды;

2) тілді зерттеуде тек адресат (тыңдаушы, қабылдаушы) факторы ғана есепке алынады;

3) тілді зерттеуде тек оның құрылымдық жағы, яғни адресант (айтушы) факторы ғана ескеріледі.

Тілді таза теориялық зерттеу барысында аталған жайлар бірдей дәрежеде орын алуы керек. Ал мәтінді өңдеуге қатысты іске қосылатын автоматты жүйелерде екінші және үшінші реттегі жайлар негізгі бағыт болып саналып жүр.

Негізінен, осы тұрғыдағы зерттеулер қиысқан мәтіндерді талдау проблемасына арналады [Гальперин, 1981]. Ал мәтінді туындату модельдерін құру жинақтау әдісін (синтездеу) қолдану арқылы іске асатынын ескерсек, қиысқан мәтіндерді жинақтау мәселесіне қатысты қазақ тілі бойынша жүргізіліп жатқан ғылыми жұмыстар санаулы-ақ. Мұндай зерттеулер нәтижесі басқа әдістер арқылы байқала бермейтін тіл табиғатының жаңа заңдылықтарын ашуға мүмкіндік жасайды.

Таза практикалық тұрғыдан қарастырғанда, мәтіндердің динамикалық туындау заңдылықтарын анықтау ісі де аса маңызды деп саналады. Әртүрлі автоматтандырылған басқару жүйелерінің даму жағдайы ЭЕМ-мен (компьютермен) "тілдесу" кезінде табиғи тіл мүмкіндіктерін пайдалану қажет болады. Сондықтан "ондай автоматтандырылған жүйелердің "адамша сөйлеу" мүмкіндігі болуы үшін оның "адамша ойлай" алатын қабілеті де болуы тиіс" деген тұжырымдарды да жиі кездестіреміз [Кузнецов, 1975, 10-15-бб.].

Егер мұндай жүйелерде "ойлау тетігі" іске асса, мәтін бойындағы неше түрлі тұрақтанған (бір үлгідегі) сипаттауларды талдау мен жинақтау жұмыстары автоматты түрде орындалар еді. Бұл жердегі "тұрақтанған сипаттауларға" (стандартное описание) мыналар жатады: мәтін ішіндегі жаңалық ашу жағдаятының баяндалу сипаты, азаматтық не әскери құжаттары мәтіндерінің ерекшелік сипаттамалары және т.б. мәтін ішінде тұрақты пішінде жиі қайталанып тұратын айтылымдар.

Мәтіннің туындау заңдылықтарын білудің келесі бір практикалық маңызды жағы – ол орта және жоғары білім беру жүйесін жаңаша ұйымдастыру мәселесі мен білім берудегі жекелік және топтық әдістер арасындағы қайшылықтарды шешуге қатысты мүмкіндіктер. Міне, оқыту процесін осы сияқты мәселелерге қатысты ұйымдастыру жұмысына жаңаша көзқарас қажет болуда. Аталған мәселелерді шешу үшін әртүрлі мәтіндердің құрамдық сипатын, олардың туындау заңдылықтарын

білу және мәтін ішіндегі негізгі және көмекші тілдік элементтерді айыра білу қажет болады. Ал бұл, бір жағынан, оқушылардың қажетті деген оқу материалын игеру ісінде, екіншіден, мамандардың жаңа оқулықтар мен оқу құралдарын құрастыру кезінде де басшылыққа алатын ұстаным.

Жоғарыда аталған мәселелер барлық түркі тілдерін, соның ішінде қазақ тілін зерттеу ісіне де тікелей қатысы бар. Әсіресе қазақ тілінің мемлекеттік дәрежесін алуы, оның халық шаруашылығын басқаруда, оқыту мен тәрбиелік жұмыстарында және қазақ халқының мәдени өмірі мен егемен елімізді қазақ тіліндегі ақпаратпен қамтамасыз ету мәселелерінде аса маңызды болмақ.

Егеменді еліміздегі қай ғылым саласын алмасақ та, оның қарқындап дамуы үшін ең алдымен халық шаруашылығын қазақ тілінде басқаратын және оқыту процесіне қажетті сапалы (әрі жедел) түрде жұмыс атқаратын автоматтандырылған жүйелерді құру мәселелері өмір талабына сай әлемдік дәрежеге көтерілуі тиіс деп санаймыз.

Әрине, мұндай мақсатқа қол жету үшін қазақ тілінің дәстүрлі грамматикаларын формалдылық тұрғысынан қайта қарастырып шығу қажет. Себебі қазақ тілінің табиғатында, басқа тілдерде де көптеп кездесетін, күрделілік пен бірыңғайлыққа жатпайтын қасиеттер және тілдік бірліктерді формалды жолмен бөліктеуде, таңбалауда, мән-мағынасын ашуда және т.б. жайттарды анықтау мәселелерінде қиындықтар көптеп кездеседі.

Бұл айтылғандар тілдік бірліктердің тұрпаты (формасы) мен мән-мағынасы арасындағы қатынастан, яғни сөзтұлға мен оның мағынасы, сөйлем тұрпаты мен оның мән-мағынасы арасынан және мәтін мен оның мазмұны арақатынастары сипаттарынан байқалады. Қазақ тілінің мұндай бірізділікке жатпайтын тұстары сөздердің не сөз тіркестерінің мағыналы құрылым жақтарында да баршылық. Міне, мұндай бірізді шешім таппайтын сипаттамалар қазақ мәтінін автоматты түрде талдау мен жинақтау жүйелеріне және білім беру саласындағы жоғарыда аталған автоматты жүйелердің дұрыс жұмыс атқаруына барынша кедергі келтіретіні сөзсіз.

Қазақ тілінің грамматикалық құрылымын дәстүрлі әдіспен зерттеу арқылы қол жеткен теориялық және практикалық нәтижелер мәтін лингвистикасы саласындағы жаңа ғылыми ізденістерге негіз болатыны белгілі.

Түркі тілдерінің синтаксисіне қатысты зерттеулерде күрделі сөйлемдердің парадигмасына күрделі синтаксистік тұтастық, абзац және төл сөз бен төлеу сөздері бар мәтін үзінділерін қосу қажеттігі ұсыныс ретінде айтылып жүр [Закиев, 1983].

Мәтіннің мазмұнын компьютер арқылы түсінуге мүмкіндік тудыратын модель жасау үшін мәтін құрамындағы негізгі тілдік бірліктердің (морфема, сөз, сөз тіркес, сөйлем және т.б.) мағыналарын формалды түрде анықтауға қатысты зерттеулер жүргізілуі керек.

Ұсынылып отырған "Қолданбалы лингвистика: формалды модельдер" атты оқу құралына автордың соңғы жылдары әл-Фараби атындағы Қазақ ұлттық университетінің филология факультетінде "Қолданбалы лингвистика" мамандығындағы қазақ бөлімінде дәріс алатын студенттерге оқыған лекцияларының материалдары негіз болып отыр. Оқу құралында ең алдымен автордың 2002 жылы орыс тілінде басылып шыққан "Основные принципы формализации содержания казахского текста" атты монографиясында қарастырылған тіл біліміндегі формалды модельдерге қатысты кейбір теориялық мәліметтер мен практикалық нәтижелер арнайы қазақ студенттеріне жеңілдетілген пішінде келтірілді. Сонымен бірге, қолданбалы лингвистика саласының белгілі ғалымы А.С.Гердтың жауапты редакторлығымен Санкт-Петербург университетінің баспасында 1996 жылы жарық көрген "Прикладное языкознание" атты оқулығындағы бірнеше ғалымдардың

мақалалары осы оқу құралы авторының ұғыну аясындағы қазақша көріністері ретінде кітаптың бірнеше тарауларынан орын алды. Анықтап айтар болсақ:

1) М.И.Откупщикова "Моделирование языка" [Прикладное..., 1996, 100-112-бб.];

2) Л.В.Бондарко "Фонетические аспекты прикладной лингвистики" [Прикладное ..., 1996, 113-141-бб.];

3) В.Д.Буторов "Моделирование синтаксиса естественного языка" [Прикладное..., 1996, 142-160-бб.];

4) В.В.Богданов "Моделирование семантики предложения" [Прикладное ..., 1996, 161-200-бб.];

5) Г.Я.Мартыненко "Методы статистического моделирования в языкознании" [Прикладное ..., 1996, 201-223-бб.];

6) В.М.Лейчик "Прикладное терминоведение и его направления" [Прикладное ..., 1996, 276-286-бб.];

7) А.С.Герд "Научно-техническая лексикография" [Прикладное ..., 1996, 287-307-бб.].

Бірақ аталған мақалалардың кейбір тұстарына автордың қалауымен қазақ студенттеріне лайықталып қажетті деген қысқартулар, түсініктемелер мен өзгерістер енгізілгендіктен оларды "Прикладное языкознание" оқулығындағы материалдардың толық түріндегі аудармасы деуге келмейді.

Автор зерттеулерінің түпкі мақсаты – қарым-қатынас кезіндегі қазақ мәтінінің мазмұнын формалдауға негіз болатын принциптерді іздеу. Ол үшін үш түрлі стиль (көркем әдебиет, публицистика, ғылыми-көпшілікке арналған әдебиет) бойынша 1500 абзацтан тұратын мәтіндер зерттеу нысаны ретінде алынды. Бұлар ең көп қолданыста болатын функционалды-мәндік сөйлеу типтері (баяндау, суреттеу, пайымдау) деп қарастырылып, қазақ тіліндегі абзац деңгейінде түзілген жиілік сөздіктерінің деректері "мәтіннің статикалық мазмұн кестесін" құруға негіз болды. Статикалық мазмұн "тірек сөздер" деп аталатын, мәтін мен абзацтарда барынша мол қолданыста болатын сөздердің айналысында шоғырланады деген ұйғарым жасалды.

Сонымен, автордың кітапта баяндалатын зерттеу идеясына бастама болған жайт: адамдардың сөйлеу қызметі кезінде әрі лексикалық, әрі синтаксистік деңгейлерде пайда болатын аса күрделі және аралас сипаттағы сабақтас (детерминді) және ықтиамалды процестер. Сөйтіп, қазақ мәтінінің семантикалық-мәнді бірліктері мен қисындасу ережелерін таңдап алу белгілі ықтималдық заңдылықпен өзгеріске ұшырап отыратын ақпаратқа негіз болады.

Оқу құралында қазақ студенттерінің назарына тұңғыш рет ұсынылып отырған тағы да бір жаңалық – көлем жағынан шағын және қазақ тілін зерттеу тарихына қатысы бар "Құдайберген Жұбанов зерттеулеріндегі модельдеу әдісінің көрінісі" атты мақала. Автор бұл мақаланы профессор Қ.Жұбановтың 105 жылдығына арналған ҚР БжҒМ Абай атындағы қазақ ұлттық университеті өткізген "Құдайберген Жұбанов тағылымы" атты республикалық ғылыми-практикалық конференцияда жасаған баяндамасынан ықшамдап алып, енгізіп отыр. Мақалада сөз болған бұл мәліметтер, ғалымның ерте кезде жүргізген зерттеулеріне қатыстылығына қарамай, әлде де қазақ тілі синтаксисі мен семантикасын модельдеу ісінде өз құндылығын жоғалтпай, заман сұранысына сай келіп тұр деуге болады.

Оқу құралында қазақ тіл біліміндегі формалды модельдерге қатысты сөз болған қысқаша мәліметтер және т.б. жайттар қазақ мәтінін автоматты түрде талдау мен жинақтау, керекті ақпаратты іздеу мен білім беру мақсаттарына өзінің оң әсерін тигізер деген сенімдеміз.

Бірінші тарау
МӘТІН – ҚАЗІРГІ ЛИНГВИСТИКАНЫҢ НЕГІЗГІ ЗЕРТТЕУ НЫСАНЫ

1.1 Мәтін ұғымы және оның анықтамалары

Мәтін зерттеу нысаны ретінде мәтінтанушылар мен стилист ғалымдардың назарына бұрыннан-ақ іліккен болатын. Мәтінтанушыларды ерте заманғы мәтіндердің әдеби-филологиялық жағы, ал стилистерді көркем әдебиет шығармаларындағы лексикалық және грамматикалық элементтерді автордың шебер қолдануы мен экспрессивтік жақтары қызықтырды деуге болады [Москальская, 1981]. Аталған зерттеу түрлерінде тілшілердің әрекеттері дайын мәтінді талдауға бағытталған болатын, ал мәтіннің туындауы мәселесінің (проблема порождения текста) тек тілге қатыссыз экстралингвистикалық жақтары ғана қызықтырды деуге болады [Супрун, 1978, 118-119-бб.].

Тіл білімінің жаңа саласы саналатын **мәтін лингвистикасы** аталған мәселені басқаша көзқарас тұрғысынан қарастырады. Е.А.Реферовскаяның айтуынша: "Предмет ее изучения – построение, порождения текста, его конструктивные элементы с точки зрения как формальной, так и содержательной структуры формирования его смысла" [Реферовская, 1983]. Табиғи тілде жазылған мәтінді ұғыну (түсіну) дегеніміз, негізінен, оның семантикасы жайлы түсінік алу, ал мәтінді туындату (генерация) – берілген семантикалық түсінік бойынша табиғи тілдегі мәтінді құру.

Мәтін әртүрлі тілдік элементтер мен олардың арасындағы қатынастардың аса күрделі және көпдеңгейлі құрылымы болғандықтан, көптеген зерттеушілер оны бірізді түсіне бермейді. Мәселен, олар мәтінді күрделі және жан-жақты құбылыс деп қарастырып, оның онтологиялық және фунционалдық қасиеттері ішіндегі ажыратылуы мен айқындалуы қиын белгісіздік, схоластикалық және сабақтастық сипаттарын ескеру қажеттігіне көңіл бөледі. Мәтін лингвистикасы саласының белгілі ғалымы И.Р.Гальперин мәтіннің мынадай анықтамасын ұсынады: "Мәтін – аяқталған, жазба құжат түрінде объектіленілген, құжаттың типіне қарай әдеби түрде өңделген сөйлеу әрекетінің шығармашылық нәтижесі. Оның құрамында әртүрлі лексикалық, грамматикалық, логикалық, стилистикалық байланыста тұратын бірнеше ерекше бірліктері белгілі мақсат пен прагматикалық ұстанымға бағытталған" [Гальперин, 1981. 4-5-бб.]. Сонымен бірге, ғалым мәтінді ақиқат шындықтың шағын бөлігінің графикалық көрінісі немесе тілдің жазба вариантының туындауы деп те түсіндіреді.

Міне, осылайша мәтіннің анықтамасына қатысты ғалымдардың әртүрлі көзқарастарын келтіруге болады. Бірақ біз оларды топтап, ғылыми бағыт түрінде беруді дұрыс деп таптық. Олар – лингвистикалық, семиотикалық, ақпараттық, құрылымдық, қолданбалы және математикалық лингвистика бағыттары.

"Мәтін" анықтамаларын қорытындылай келе былай тұжырымдауға болады. Мәтін – белгілі құрылымдық заңдылықпен жинақталған, фразадан тыс тұтастық

құрап, өзара қатынаста тұратын сөйлемдер жиынтығының іріленген бірлігі (абзацтар, параграфтар, бөлімдер, тараулар және т.б.). Бұлар коммуникация мақсатына сай белгілі аяқталған мазмұнды тарату үшін қызмет етеді [Гальперин, 1974].

Егер мәтінді тек сөйлем деңгейінде қарастыратын болсақ, онда оны бөлектенген (дискретті) сөйлемдер топтарынан тұрады деп пайымдауға болады. Ал мәтіннің сызықтық тізбек құрайтын қасиеті, сөйлемнің монологтық айтылымдарда ең кіші коммуникативтік бірлік екендігі айқындала түседі [Зарубина, 1981, 22-б.].

Мәтіннің тұтас және тұйық коммуникативтік бірлік болуына байланысты оның негізгі құрастырушы белгісі – семантикалық қиысушылық (связность) пен семантикаға қатысы бар грамматикалық және лексикалық формалдылық.

Ақиқат шындықта тілге тікелей қатысы жоқ әрекет, оқиға, жағдаят (ситуация), өзара тәуелділікті және т.б. тіл арқылы хабарлау үшін мәтін қиысқан сипатта болуы керек. Сондықтан қиысқан мәтіннің лексикалық құрамының өзіне ғана тән ерекшеліктері бар. Мәтіннің өн бойында мазмұнды тұрақтандыру қажеттігінен туындайтын "тақырыптық сөздер" тобы үнемі даму сипатында болады. Мәтін тұтастығын сақтайтын деиктикалық элементтер мен синтаксистік байланыстардан басқа, қисынды мәтінге шақ формаларының да бірден-бір қатысы бар. Олар баяндаудағы әрекет субъектісі мен қалыптық сипаттарды мезгілдік өлшемге тәуелді етіп отырады.

Мәтіннің келесі бір сипаты – мәтін бірліктерінің өзара байланыстылығының (қиысу мүмкіндігінің) бір мәнді болмауы. Мұндай сипат кез келген ақпараттық процестің ішкі заңдылығына байланысты және оны автордың мәтін құру жағдаятынан байқауға мүмкіндік туады. Қалам иесі өзінің әлемтану мүмкіншілігіне қарай, яғни сыртқы әлем жайлы білім қорына, лингвистикалық дәрежесіне және эмоционалдық қалпына қарай ақиқат шындықты суреттей алады [Новиков, 1981, 48-56-бб.].

Мәтінді оның семантикасы тұрғысынан **дәстүрлі семантикалы мәтін** және **өз семантикасын өзі туындататын мәтін** деп екі топқа бөліп қарастыруға болады [Шрейдер, 1976, 153-169-бб.].

Бірінші топқа суреттеу типті мәтіндер жатады. Олардың референттері алдын ала анық (немесе автор санасында сезілетін) болуы керек. Ал өз семантикасын өзі тудыратын мәтіндерде референттер жаңадан туындайды. Бұған математикалық конструкцияларды суреттеуге, жобалауға қатысты мәтіндерді және көркем әдеби шығармалар сипатына тән келетін барлық мәтіндерді жатқызуға болады.

Мәтін лингвистикасы тұрғысынан мазмұн түрлері де жіктеліп сипатталады: тілдік және мағыналық; ұғымдық және қорытындылау [Реферовская, 1989].

Сонымен, нақты бір мәтін мазмұнының оны құрайтын бөліктер мазмұнына қатысы бар (бөлім, тарау, мәтін кесінділері және фразадан тыс бөлік) деп саналады. Бірақ мәтіннің мән-мағынасы (смысл текста) оның құрамындағы сөйлемдер мағынасының жиынтығынан тұрмайтыны белгілі, себебі әр сөйлем өзінше "көрінбейтін" және лингвистикалық тұрғыда өрнектелмейтін "ішкі мәтіннен" тұрады. Мәтіннің толық мән-мағынасы оның мазмұнын жалпылап, қорытындылап, түйін жасау арқылы іске асады. Мәтін негізінде жатқан жалпы пікір – мәтіннің жалпы мәні болып саналады.

Сөйтіп, мәтіннің лингвистикалық құрылымын зерттеу оның негізгі бірліктерін анықтау мен оларды бөліп алудан басталады. Мұндай жағдайда аса назар аударатын мәселе сөйлемнің тілдік және коммуникативтік құрылымы. Осымен бірге, сөйлемнен үлкен бөліктер және ең соңында мәтіннің өзі тұтас түрде қарастырылуы қажет.

Мәтінді коммуникативтік бөлшектеу, негізінен, психолингвистикалық категорияға қатысты деп саналады. Яғни ол ақиқат әлем құбылысының объективті және нақты қатынастарымен бірге, оларға деген адам баласының жеке көзқарасынан тұрады деп ұғыну қажет.

Сонымен, "мәтін" ұғымын талдай келе, кез келген мәтінге тән мынадай қасиеттерді атауға болады:

1) мәтіннің бір ғана мазмұнға ие болуы;

2) мәтінді құрайтын белгілі бірліктердің болуы;

3) мәтін бірліктері арасындағы грамматикалық және логикалық байланыстың болуы;

4) мәтіннің нақты модальдылық пен прагматикалық ұстанымға ие болуы;

5) мәтін бірліктері арасындағы байланыс сипатының тұрақсыздығы.

Мәтін құрылымын тануды мақсат еткен кез келген зерттеу үш түрлі мәселені шешуді мақсат етеді:

- мәтіннің негізгі бірліктерін анықтау;
- негізгі бірліктерді синтагмалау ережелерін айқындау;
- негізгі бірліктер негізінде іріленген мәтін бірлігін жинақтау ережелерін табу.

Зерттеуші ғалымдардың көпшілігінің пікірінше, сөйлемнің мағынасы оның құрамындағы сөздер мағыналарының "қосындысы" болмайтындығы тәріздес, мәтін мазмұнын да оны құрайтын сөйлемдер мағынасының жиынтығы деуге болмайды.

Сонда мәтіннің негізгі семантикалық компоненті дегеніміз не? Міне осы мәселеге қатысты жайтттар әрі қарайғы баяндауларымызда сөз болады.

1.2 Мәтін және тілдік таңба

Таңбалар мен олардың жүйесінің жалпылама қасиеттерін танып-білумен семиотика ғылымы шұғылданатыны белгілі. Солай бола тұра, қазіргі кездегі басқа ғылым салалары да, мысалы, философия, логика, кибернетика, лингвистика, психология және т.б. таңбалар жүйесіне арнайы бөлім арнап, айрықша көңіл бөлуде. Осының нәтижесінде семантика ғылымы пәнаралық сипат алып, көптеген қоғамдық және жаратылыстану ғылымдарының даму барысына өз ықпалын тигізіп келеді.

Таңбалар жүйесін зерттеудегі семиотикалық бағытқа сәйкес коммуникация мен тіларалық қатынасты жасанды таңбалар жүйесімен салыстырғанда табиғи тілдің біріншілік сипаты ескерілуі қажет.

Қоғам өмірінің күрделілігі, сана-сезімнің баюы мен даму сипаттары адамдар арасындағы қарым-қатынас түрлеріне де өз әсерін тигізбей қоймайды. Мұндай әсерді табиғи және жасанды тілдердің дамуы мен күрделену сипатынан аңғаруға болады [Салмина, 1988, 15-б.]. Ал бұл факт, таңба мен таңба қызметінің әлеуметтік прогресс пен адам баласының өмір сүруінің маңызды элементі екенін байқатады. Электронды-есептеу техникасының көмегімен ақпаратты қадағалау, жинақтау мен өңдеу процестерінде таңбалар мен олардың жүйелерін саналы түрде құрастыру қажеттігі туындап отыр. Адам өмірінің коммуникативтік қызметінде табиғи тілдің әмбебаптық және қисындылық сипатына қарамай, аталған амалдарды ақпаратқа қатысты іске асыруда табиғи тіл жасанды тілдің мүмкіндігінен асып кете алмай жүр. Бұған дәлел ретінде өмірде іске асып жүрген таңбалар жүйесі заңдылықтарының негізінде құрылған неше түрлі автоматтанған ақпараттық жүйелер (ААЖ) мен ақпараттық-іздеу жүйелерін (АІЖ) атауға болар еді. Соңғы кездегі тіл біліміндегі зерттеу нысанының негізгілерінің біреуі ретінде "мәтіннің" алынуы әлеуметтік

қажеттілікпен бірге, тілтану ілімінің ішкі дамуының қисындылығынан деуге болады. Ақпараттық ғылым саласының белгілі ғалымы А.П.Ершов осы айтылғанға қатысты былай дейді: "... Если расчетные инженерно-научные и экономические применения ЭВМ предопределили облик вычислительного дела в 50-е и 60-е годы, то сейчас центр научной проблематики смещается в сторону обработки текстовой информации в широком смысле, а также массового применения ЭВМ" [Ершов, 1979].

Ғылыми-зерттеу ісінің жаңа бағытқа қарай ауысуына себеп болып отырған, біріншіден, ЭЕМ арқылы ақпарат көлемін "қысу" (жинақтау), оны ұзақ мерзімге сақтау мен қайта өндіру немесе баспа ақпараттарын тиісті белгілеріне қарай іздеп табу. Екіншіден, мұндай зерттеу әдісінің жаңа бағытқа ауысуы ғылыми мақалалардан аннотация, реферат алу, шет тіліндегі әдебиеттерді аудару және оқу орындарында компьютерлік білім беру және т.б. мәселелерімен байланысты деуге болады.

Сонымен, тіл қызметі жаңа сипаттағы функцияларға ие бола бастады. Бірақ таңба мен таңбалар жүйесін құрастыруға қатысты мәселелерді шешу семиотиканың негізгі ұғымдарын жете ескеруді және тіл қызметінің принциптерін толық түсіне білуді қажет етеді.

"Таңба" ұғымының табиғатына терең үңілмей-ақ, мына анықтамаға назар аударайық: "Таңба – басқа бір заттың (құбылыс, оқиға) орнына жүретін белгі. Ондай белгі арқылы заттың (құбылыстың, оқиғаның) қасиеттері мен қатынастары негізінде ақпаратты қабылдау, сақтау, өңдеу және тарату іске асады" [ФЭС, 1983].

Қазіргі кезде "таңба мағынасының" проблемасы әр білім саласының көптеген өкілдерін қызықтыруда. Таңбаның материалдық қасиетіне қарағанда, оның "мағынасы" айрықша түрдегі функционалдық қасиеті деуге болады, яғни ол таңба қызметінің идеалды қасиеті болып саналады. Таңба мағынасының табиғатын бұлайша түсіну оның тағы бір қырын (таңба антиномиясын) айқындау үшін қажет. Дәлірек айтсақ, мағына таңбаның өн бойында ма, әлде одан тыс жатыр ма деген сауалды түсіндіруге мүмкіндігі бар.

Көптеген ғалымдардың мәтін мағынасының табиғатына қатысты көзқарастарын былайша қорытындылауды жөн көрдік. Таңбаның мағынасы қоғамдық танымдық әрекет кезінде пайда болады және "өмір сүреді". Бұл жағдай адам баласы мен ақиқат шындықтың басқа нысандары арасында пайда болатын күрделі байланысының санада бейнеленуі арқылы пайда болатын әсері.

Таңбаның келесі қасиеті – ол таңбалар жүйесіндегі таңба тұрпатының шарттылығы, яғни оның таңбаланатын затқа қарағанда шартты және еркін белгіленуі. Мұндай тұжырымдау таңбаның материалды зат орнына жүрмейтіндігінен, оның кейбір қызметін ғана атқара алатындығынан туындайды.

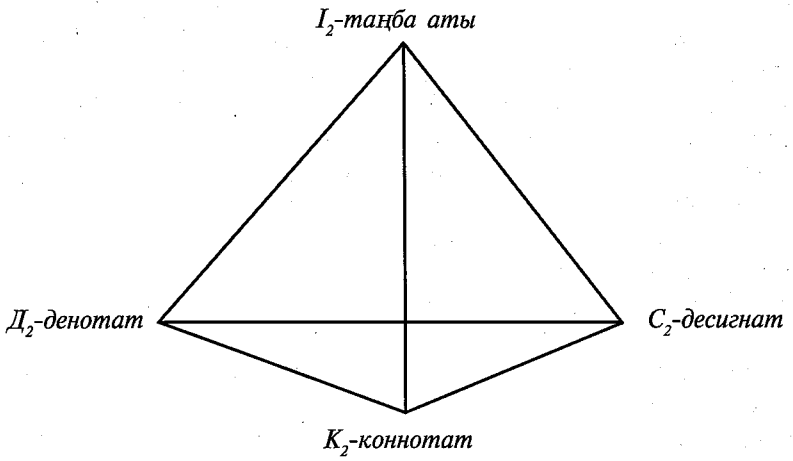
Тілдік таңбаға қатысты әрқилы семантикалық сипаттаулардың (логикалық, прагматикалық, лингвистикалық) ішінен, біздің ойымызша, айырықша назар аударуға тұратыны Р.Г.Пиотровскийдің таңбалар жүйесіне байланысты берген инженерлік-лингвистикалық түсінігі. Бұл теория бойынша, анықтаушы референтте дербес сипаттағы кем дегенде тәуелсіз екі компоненттің – денотат пен десигнаттың қатысы бар деп тұжырымдалады. Бұлардың біріншісі "зат атына" немесе таңба атына қосылып, референттің толық күйінің адамның сезім мүшесінде шағылысуы болып саналады. Ал "зат аты" деп отырғанымыз – сыртқы әлемнің қайсыбір жекеленген нысаны. Екінші компонент – десигнат (интенсионал, концепт, сигнификат, мән) нақты "зат аты" (таңба аты) арқылы белгіленетін және концептіні айқындайтын мағыналық бірлік, яғни бір топқа қатысты заттар жиынтығының жалпы мән-маңызы.

Осы теорияны әрі қарай дамыта отырып, Р.Г.Пиотровский таңба құрылымына үшінші компонентті – коннотатты енгізеді. Бұл компонент жоғарыда аталған денотат

пен десигнатқа үстеме ретінде екінші мағыналық реңкті және бағалауды немесе эмоционалды-бағалық аспектіні үстемелейді.

Р.Г.Пиотровский өзінің таңба теориясына қатысты зерттеулерінде "таңбаны" даму үстіндегі (динамикалық) жүйенің элементі ретінде қарастырады.

Сонымен, семиотикалық зерттеулерді тұңғыш рет өмірге әкелген Г.Фрегенің "таңбалық жағдаятының" (знаковая ситуация) үш бұрыштық берілісінен (таңба аты, денотат, мағына) кейін, Р.Г.Пиотровский мен оның шәкірттерінің зерттеулерінде таңба аты мен референт арасындағы қатынас *1-суреттегідей* көрініс тапты:



1-сурет. Сөз-таңбаның сызбасы

Қолданбалы тіл білімі саласының белгілі ғалымы Р.Г.Пиотровский аталған қатынасты нақтылай түсіп, яғни таңбаны көп компонентті психикалық негіздегі беріліс (психическая сущность) дей отырып, оны алты компоненттік құрылымға жатқызады.

Олар:

1) референт (зат);
2) денотат (зат бейнесі);
3) десигнат (концепт, мән, ұғым);
4) сигналды референт;
5) коннотат (белгінің эмоционалды-бағалық аспекті);
6) таңба аты, таңбалаушы.

Біздің әрі қарайғы баяндауымызда сөз-таңбаға қатысты мына ұғымдар қолданылуы мүмкін:

1) референт – нақты материалдық зат немесе солардың тобы (класы). Бұлар адамның сана-сезімімен, қиялымен бөлініп алынған дерексіз ұғым немесе жоқ (болмайтын, кездеспейтін) нысан, заттың не оның топтары сипаты мен ұғымдарының аралық қатынас түрлері де болуы мүмкін.

2) денотат – референттің адам санасындағы бейнеленуі;

3) десигнат – референттің ең маңызды қасиеттерінің адам санасындағы бейнеленуі.

Сонымен, жоғарыда көрсетілген *1-суреттегі* сөз-таңба сызбаны ұстаным ете отыра, таңба мағынасы деп таңба аты мен осы таңба компоненттері арасындағы қатынастардың адам сана-сезіміндегі жиынтық көрінісі деуге болады.

Электронды-есептеу машинасында (компьютерде) "сөз-таңбаны" қолдану кезінде тілдік таңба сипатының біртіндеп қарапайымдану сипаты байқалады. Ол, ең алдымен, тілдік таңбаның психикалық дәрежесінің біртіндеп төмендеуінен аңғарылады [Пиотровский, Шингарева, 1979, 186-б.].

Компьютерде іске асып жүрген тілдік (жазба және сөйлеу) модельдерде таңба аты (имя) мен денотаттың кілттерін қолданатын сөздер таңбасының моделі пайдаланылады. Ал біздің зерттеу модельдерімізде аталған кодтардан басқа сөздердің семантикалық ішкі топтары түріндегі десигнат коды да ескеріледі.

Енді келесі баяндауларымызда тілдік бірліктерді динамикалық сипатта қарастыруды (сөйлеу қызметін) бағыт-бағдар етіп алып, әр түрлі тілдік таңбалардың типтері үшін Г.Фрегенің семиотикалық сызбасынан Р.Г.Пиотровскийдің таңба моделінің динамикалық жүйесіне өтудің ретін қарастырамыз.

Тілдік таңбаны осылайша функционалды бағытта қарастыру қабылдаушыға таңбаны бейне, мазмұн және түсінік тұрғысынан қарастыруға мүмкіндік туғызады.

Қорыта келе, таңбаларға тән жалпы қасиеттерден тілдік таңбаға қатыстылары деп мыналарды атауға болады:

1) тілдік таңба сезім мүшелерінің бірі арқылы білінетін сипатта болады;

2) бір нәрсе жайын хабарлайды, коммуникативтік қызмет атқарады;

3) белгілі бір мағынаға, мазмұнға бағытталады, онымен шартты қатынаста болады;

4) өзі білдіретін мағынамен қатынасы еркіндік сипатта болады, олардың арасында табиғи байланыс болмайды;

5) әр таңба өз жүйесіндегі басқа таңбалардан бір немесе бірнеше айырым белгілері арқылы жекеленіп танылады [Қордабаев, Қалиев, 2004, 76-б.].

Демек, тілдік жүйе мейлінше күрделі, көп сатылы, көп қабатты, көп элементті болып келеді. Оған сөздерді, сөз тіркестерін, сөйлемдерді, абзацтарды және одан да үлкен мәтін бөліктерін жатқызуға болады.

1.3 Сөз тіркесі – абзац бірлігі

Жазба тілді қабылдауға қатысты психологиялық зерттеулердің деректеріне сүйенсек, жазба тіл ағымы мағыналық байланыста тұратын жекеленген тілдік бөліктерге бөлшектеніп, ал олар оқырманға сөз тіркес деңгейінде қабылданады. Мұндай пікір лингвистік және әдістемелік материалдардың зерттеу нәтижелері арқылы да қолдау тауып жүр. Осындай қорытындыны айтылымның туындау процесін ритмика-просодиялық талдаудың нәтижелері де мақұлдайды.

Ғалымдардың байқауынша, егер сөйлем деңгейінде ең кіші бірлік ретінде **сөз** алынатын болса, абзац деңгейінде оның құрамындағы сөздерге барабар ең кіші әрі күрделі номинативті бірлік рөлін **сөз тіркес** атқарады [Долгова, 1980, 148-б.]. Мұндай синтагма типті интонациялды-мәндік бірліктерді бөліктеудің қажеттігін жарыса (синхронды) аударма жасайтын мамандар да қолдап жүр [Ширяев, 1979, 109-б.].

Сөз тіркестерін, мәтіннің мүмкін болатын семантикалық компоненті деп есептеп, жеке талдайтын болсақ, онда олар сөздерге қарағанда ақиқат шындықтың элементтері арасындағы қатынасты білдіре алады. Адамның сана-сезімінде ондай қатынас жалпылама сипат күйінде көрініс табады. Бірақ сөйлеммен салыстырғанда

сөз тіркесінің мағынасы хабарламаның аяқталған түріне және ондағы орын алатын ақпараттың объективті-модальдық әрі мезгілдік сипатына сай келе бермейді.

Ақиқат шындықтың нысандары өзінше бөлек және таза күйінде кездеспейді деуге болады. Олардың әрбіреуі жағдаятқа қатысты басқа бір нысандармен салыстырғанда қайсыбір сапалық және сандық сипаттармен ілесе жүруімен ерекшелінеді. Жағдаят құрастырушыларының осындай ерекшеліктері мәтін ішінде белгілі бір **сөз тіркестері** арқылы нақтыланады. Ғылыми әдебиеттерде сөз тіркестерінің мұндай түрлері *атрибутивті, объективті, предикативті, адвербальды* және т.б. деп аталып жүр.

Адам өзінің балалық шағы кезінен бастап, қардың – *ақ,* аспанның – *көк* екендігін, құстың – *ұшып,* жыланның – *жорғалайтындығын* және т.б. есінде қалдыру үшін: *"ақ қар", "көк аспан", "құс ұшады", "жылан жорғалайды"* және т.б. сөз тіркестерін пайдаланады. Сол сияқты *"әкесінің аты", "бөлменің есігі"* болатындығы, *"тәртіпті болу", "тентек болу", "көп жылау", "тамақ ішу"* және т.б. әрекеттерді де аталған сөз тіркестері арқылы есінде сақтауға тырысады. Өмір сүру мен тілді үйрену (үйрету) процестерінде де ең жиі қолданыстағы сөз тіркестері адам есінде тез сақталып, олардың саны әрдайым ұлғайып, адам санасында "сөз тіркестері қорына" айналады. Мұндай процесс, негізінен, бала кезден бастап қалыптасады деп саналады. Оған себеп болатын: үлкенге еліктеу жолдары, ата-ананың (үлкеннің) бала тілін қалыптастырудағы белсенді әсері және жас баланың өз ойын жинақтап, қорыту мүмкіндігі т.б. [Тер-Минасова, 1981, 27-б.]. Сол сияқты, жазушылар да өздерінің шығармашылық әрекетінде негізінен есінде қалған немесе қажеттікке сай өзіне дейінгі қоғамда орын алған сөз тіркестерін жағдайға қарай қолдануға тырысады, қажетті мәнмәтіндерде (контексте) қайталайды [Тер-Минасова, 1981, 40-б.]. Басқаша айтқанда, мәтін құрылымының элементтері сөйлем де, сөз де емес, онда зат пен оған тән белгі-қасиеттер тілдік мүмкіндіктер арқылы көрініс табады екен. Яғни көп жағдайда, мәтін абзацтары әрқилы сөз тіркестер типтерімен беріледі деуге болады. Әрине, мұндай сөз тіркестерінің құрамы көптеген факторларға байланысты автор тілінің шеберлігіне де тән қасиеттерге ие. Мәселен, естияр адамдар үшін сөз тіркестерінің құрамы көбінде оның мамандығына, әлеуметтік жағдайына, қоршаған ортасына байланысты болып келеді.

Құрамы мен оған енетін бірліктердің саны жағынан **сөз тіркестері** мәтін ішінде түрлі-түрлі болып кездеседі. Мысалы, олар *жаңа үй, сенімді ақтау, ғылыми дәреже* сияқты жай тіркестермен қатар *көп оқысаң көп білесің, жаңа заманның жалынды жастары* және т.б. күрделі сөз тіркестері болуы да мүмкін. Солай бола тұра, адамдарға үйреншікті, айтылуға жеңіл сөз тіркестердің типтері құрам жағынан көп жағдайда екі сөзден тұратыны аян. Олардың біреуі заттың атауы болса, екіншісі оның белгісін білдіреді. Яғни жоғарыда келтірілген мысалдардағыдай: *ақ қар, жақсы кісі, алтын сағат* және т.т. Мұндай тіркестер **жай тіркес** немесе **элементар тіркес (ЭТ)** деп аталып жүр. Сөзді сөзге тіркеу тәсілдеріне қарай **элементар тіркестер** ғылыми әдебиеттерде үш түрлі типке бөлініп, ажыратылады.

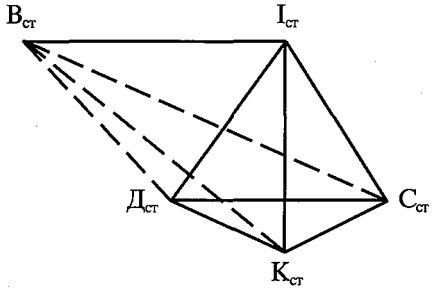
Бірінші **ЭТ** типке зат пен оның жай белгісі арасындағы қатынасты білдіретін:

а) атрибутивті тіркестер типі: **"сын есім+зат есім"** (немесе керісінше);

ә) субстантивті тіркестер типі: **"зат есім+зат есім"**.

Екінші **ЭТ** типі зат пен оның қимылдық (әрекеттік) белгісі араларындағы қатынасты білдіреді – **"зат есім+етістік"** (немесе керісінше). Ал үшінші **ЭТ** типіне әрекет белгісі мен жай белгі арасындағы қатынасты білдіретін элементар сөз тіркестерін, яғни **"етістік (есімше)+үстеу"** (немесе керісінше) жатқызуға болады.

Біздің сипаттауымызша элементар сөз тіркесінің (**СТ**) семиотикалық құрылымы *2-суреттегі* сызба түрінде көрініс табады.



2-сурет. Элементар сөз тіркестерінің (СТ) семиотикалық құрылымы

Осы көріністен байқайтынымыз элементар сөз тіркестері бес компонентті қатынастар негізінде сипатталады.

Олар:

- $І_{ст}$ – таңба аты (екі тізбекті атау);
- $Д_{ст}$ – денотат;
- $С_{ст}$ – десигнат;
- $К_{ст}$ – коннотат;
- $В_{ст}$ – екі элементтен тұратын сөз тіркесі арасындағы синтаксистік қатынас.

Сөз тіркестері де сөздер сияқты мағынаға ие болады десек, "сөз тіркесі" мағынасы адам санасында жинақталған "сөз тіркесі" атының, осы "сөз тіркесі-таңбаны" анықтайтын барлық компоненттерге қатынасы деуге болады.

Сөз тіркесі моделінде таңбаның компьютердегі қызметі таңба атының, денотат пен десигнаттың кодтары және синтаксистік қатынас коды арқылы іске асады.

1.4 Сөйлем – тілдік таңба және мәтін бірлігі

Ғалымдардың сөйлем семантикасына қызығушылығы өткен ғасырдың 60-жылдарының ортасынан басталады деуге болады. Бұл проблема әртүрлі көзқараста: лексикологиялық, жағдаятқа қатысты мағына тұрғысынан, логикалық және арнайы лингвистикалық тұрғыда қарастырылды [Арутюнова, 1976.]. Сөйлем семантикасы қандай ғылыми бағытта қарастырылса да, зерттеудің негізгі мақсаты айтылым мен экстралингвистикалық жағдаят немесе оқиға араларындағы қатыстықты анықтау болып табылады.

Тілшілер соңғы кезге дейін сөйлем мағынасын мәнмәтіннің қатысынсыз, бөлек қарастырып келді. Мұндай әдістемелік бағыт синтаксистік парадигматика, деривация және трансформация әдістерінде де орын алады. Сөйлемді бұлайша зерттеудің негізгі себебі оның пішіндік берілуіне қарағанда мазмұн тұрғысына көбірек көңіл бөлінуінде деп түсінген жөн.

Қазіргі кезде сөйлем семантикасын тануда мәтінді мәнмәтінді-қиысу синтаксис ережелері арқылы сипаттау жиі қалыптасуда.

Сөйлем семантикасын зерттеуде көп нәрсе оның негізі ретінде қандай тұжырымдаманы ұстаным етіп алуға байланысты. Осы айтылғанға байланысты ең жиі ұстанымдар мыналар:

1) тілден тысқары жағдаят – референттік тұжырымдама;

2) ойлау құрылымының заттық-атрибутты пайымдау түрі немесе пропозионалды функция түрі – логикалық тұжырымдама;

3) Verbum finitum және оның грамматикалық категориясы – морфологиялық тұжырымдама.

Осы аталғандар ішінен сөйлем семантикасының негізі болатын референттік тұжырымдамаға жеке тоқталайық.

Тіл бірліктерінің таңбалық табиғатына сай сөйлемнің мән-мағынасы былайша түсінік табады. Сөйлемнің мән-мағынасы дегеніміз – қайсыбір қарапайым жағдаят пен оқиға күйіндегі ақиқат шындық бөлігінің адам сана-сезіміндегі бейнеленуі [Гак, 1976, 22-26-бб.]. Осындай қарапайым жағдаятты психикалық жағдаяттан және мән-мағына бойындағы денотативті мен десигнативті құрастырушылардан ажырата білуіміз қажет. Бұл жердегі **сөйлемнің денотаты** дегеніміз адам санасында бейнелетін қайсыбір нақты **"жағдаят"**, ал **десигнат** – біртекті жағдайлар жиындары ішіндегі белгілі бір **"жағдаят"** топтамаларының адам санасында қалыптасатын көрінісі екен [Гак, 1977, 232-б.].

Шығарма авторы нақты сөйлемді жазуда тек қана белгілі бір оқиғаны баяндаудан басқа, ақиқат шындықтың біртекті жағдайлары ішінде кездесетін жалпы "бірдемені" де (нечто) білдіргісі келетіні белгілі. Сонымен бірге жазушы (сөйлеуші) сөйлемді мазмұндық аспектіде қарастырғанда, неше түрлі эмоционалды, экспрессивті және бағалаушы реңктерді де білдіруге ынталанады.

Л.М.Васильевтің ұйғаруы бойынша, айтылым түріндегі сөйлем мазмұны – номинативті, модальды және коммуникативті аспектілерді қоса білдіретін сөйлеудің мән-мағынасы. Мұндағы сөйлемнің модальды аспектілік жағы модальдық пен темпоральдықтың (мезгілдік, шақтық) семантикалық категорияларын нақты түрде іске асыру ретінде қарастырылады.

Егер сөйлемнің модальдық және номинативтік аспектілерін бірге қарастыратын болсақ, онда біз сөйлемнің семантикалық құрылымын, денотативті және сигнификативті семантикасын анықтай аламыз. Ал енді, тек оның коммуникативті жағы ғана қарастырылса, онда сөйлемнің деиктикалық табиғатын айқындауға мүмкіндік туады [Васильев, 1988, 5-10-бб.].

Тілдік бірліктерді мазмұндық тұрғыда зерттеу үшін мынадай негізгі үш фактор біртұтас түрдегі қалыпта ескерілуі қажет. Бұл факторлардың семантиканы қалыптастыруға қатыстылары: тілдік бірліктердің тұрпаты, денотат және сигнификат [Плотников, 1984, 89-б.].

Ақиқат шындықтың кейбір ареалдарының адам сана-сезімінде бейнеленуіне қарай, айтылым аясындағы семантикалық элементтерді үш типке бөліп қарастыруға болады. Біріншісі – адамнан тыс жатқан объективті әлемнің жағдайларын бейнелейтін семантика-синтаксистік тип; екіншісі – адамның сөйлеу процесі мен психикалық қызметін бейнелейтін интенсионалды-семантикалық тип; үшіншісі – адамзаттың ойлау қызметін (қабілетін) бейнелейтін логика-семантикалық тип [Чахоян, 1988, 11-б.].

Сөйлемнің таңбалық құрылымындағы арнайы көңіл бөлуге тұратын жағы – прагматикалық аспект. Мұны сөздің таңбалық құрылымындағы коннотативті аспектіден кең мағынада түсінген жөн. Себебі, ол басқа да ақпараттық және коммуникативтік құрастырушыларды бірге қамти алады.

Мәтіннің негізгі семантикалық компоненттерінің бірі ретінде **"сөйлем"** бірлігінің саналуы ғалымдар арасында бірдей қолдау таппай жүргені белгілі. Кейбір ғалымдар

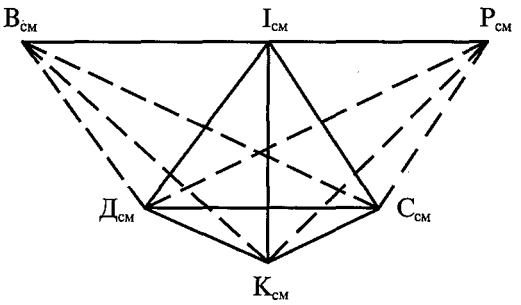
сөйлем жеке тұрып, яғни мәтіннен бөлек тұрып дербес мағынаға ие болмайтындығын баса айтады. Сөйлем мән-мағынасының ақиқат шындықтың белгілі бір заттық сипаттағы бөлігінің адам санасындағы бейнеленуі екендігін ескере отырып, оны мәтіннің негізгі мағыналық бірлігі деп есептейді. Ал тіл бірліктерінің негізі ретінде көптеген жағдайда сөйлемнің алынуына қарамай, кейбір ғалымдар оны мәтін бірліктерінің де негізгі бірлігі деген пікірмен келісе қоймайды. Тағы бір зерттеушілер **сөйлемді** тілдің ең үлкен бірлігі және сөйлеу тілінің ең кіші бірлігі деп есептейді. Сонымен бірге олар сөйлемнің әрбіреуінің нақты мазмұнының ерекшелік және жекелік сипаты негізінде **сөйлемді** тілдің айрықша бірлігі деп санауды ұсынады. Сондықтан **сөйлемнің семантикасын** зерттеу үшін оның өзіне ғана тән айрықша әдістемелік бағыт қажет деп саналады.

Мәтін лингвистикасы саласындағы көптеген зерттеу нәтижелері бойынша **сөйлемнің** дербес бірлік еместігі анықталды. Яғни оның жоғары деңгейдегі тілдік бірліктер: абзацқа, фразадан жоғары тұтастыққа немесе барлық мәтіндерге тәуелді екендігі дәлелденіп отыр. Кейбір ғалымдар тіпті мәтін аясының өзінде де, **сөйлем** туралы емес, **жалған сөйлем** (псевдопредложение) жөнінде сөз етуді дұрыс деп санайды [Звегинцев, 1976, 185-188-бб.].

Мұндай пайымдаулар сөйлемнің мәнмәтіндік тәуелділік сипатын дәлелдей түседі, яғни бұл жағдай сөйлемнің өз аясынан тыс жатқан көптеген факторларға байланысты болуының айғағы.

Сөз тіркестеріндегі сөздер қызметі сияқты сөйлемдер құрамындағы сөздер де бір-бірімен, белгілі дәрежеде, синтаксистік қатынаста болады.

Біздің пайымдауымызша, **сөйлем-таңбаның семиотикалық құрылымы** *3-суреттегідей* көрініс табуы керек.



3-сурет. Сөйлем-таңбаның семиотикалық құрылымы

Мұндағы:
- $I_{см}$ – сөйлем-таңба атауларының тізбегі;
- $Д_{см}$ – сөйлемнің денотаты;
- $С_{см}$ – сөйлемнің десигнаты;
- $К_{см}$ – сөйлемнің коннотаты;
- $В_{см}$ – сөйлемнің синтаксистік аспекті;
- $Р_{см}$ – сөйлемнің ақпараттық аспекті.

Жоғарыда пайымдалған сөздің, сөз тіркесінің және сөйлемнің таңбалық сипаттарын салыстыра келе байқайтынымыз – тілдік таңбаның деңгейі жоғарылаған сайын, олардың таңбалық сипатының күрделене түсуі.

Компьютер жадындағы сөйлем-таңба қызметіндегі сөйлемнің таңба атауларының кодтары, сөздер денотаттарының жиынтығы ретіндегі сөйлемнің

денотат және десигнат кодтары, сөйлемнің синтаксистік және ақпараттық аспектілерінің кодтары пайдаланылады. Аталғандардың соңғы екеуі біздің зерттеуімізде СЕМСИНТ жасанды тілінде жазылған сөйлемнің семантика-синтаксистік өрнегін құру үлгісінде (моделінде) жүзеге асты.

1.5 Абзац – мәтіннің негізгі семантикалық бөлігі және таңба

Мәтін ішіндегі абзацтар мәселесі ғалымдар арасында әлі де бірізді көзқарасқа, анықтамаға ие болмай келеді. Қолданбалы лингвистика саласының мамандары абзацты мәтін бірлігі ретінде не композициялық, не логикалық, не семантикалық бірлікке жатқызып жүр.

Сөйлемдерді аса ірі мәтін бірліктеріне жинақтайтын белгілердің – сызбалық, синтаксистік және семантика-синтаксистік түрлерін айрықша атап кетуге болады. Қазіргі кезде осы аталғандардың ішінен зерттеу әдістері ретінде семантика-синтаксистік бағыт ең жиі қолданыс тауып жүр. Сөйлемдер тізбегін бұл әдіспен жинақтау (бөліктеу) – "күрделі синтаксистік тұтастық", "фразадан жоғары бірлік (единство)", "прозалық (прозаическая) строфа" және т.б. ұғымдар арқылы аталып жүр. Мұндай ұғымдардың жиырмадан астам анықтамалары бар деуге болады. Олардың барлығының да сөйлемдерді топтауға, жинақтауға қоятын талаптары ұқсас, яғни мәтіннің мұндай бөлігіндегі сөйлемдер арасында мағыналық және синтаксистік байланыстың болуы. Бірақ мұндай анықтамаларда мәтінді бөліктеудің формалды жақтары әрқилы түсіндіріледі. Мысалы, кейбір авторлар үшін "күрделі синтаксистік тұтастық" абзацпен сәйкес келетін болса, басқаларында ол не абзацтан кіші немесе бірнеше абзацтарды біріктіріп қарастырумен парапар болады.

Солай бола тұра, зерттеушілердің айтуынша, сөйлемдерді формалды топтау автордың жазу әрекетінің ырқына бағынбайды. Мәселен, психологтардың тәжірибелеріне сүйенсек, мәтінді формалды бөліктеу негізінде оның сыртқы белгісінен гөрі, мән-мағыналық мазмұны көбірек орын алады. Жазушының шығармашылық ой дамыту процесінде мәтіннің әр бөлігі өзінше бөлек немесе салыстырмалы жақындықтағы мазмұндық тұтастыққа ие болады. Бұл туралы тәжірибеге қатысқан қалам иелерінің "өз ойларын топтайтыны", "тақырыпқа қарай мазмұндық тақырыпшаларға бөліктейтіндігі" дәлел бола алады. Шындығында, мұндай әрекет мән-мағыналық топтау деп аталады. Тәжірибе нәтижелері бойынша осылайша бөліктеу, көбінде, абзацпен сәйкес келетіні анықталды [Бондарко, Шрейдер, 1978].

Мәтінді абзацтарға бөліктеу, ең алдымен, мәтін мазмұнын жақсы меңгеруге және есте сақтауға ықпал етеді. Екіншіден, абзац бөліктері ойдың логикалық бірлігін, тұрақтылығы мен түйіндеу қалпын сақтайды [Колшанский, 1983, 29-33-бб.].

Сонымен, **"абзац"** дегеніміз, шындығында, сөйлеу тәжірибемізде бар ақиқат мәтін бірлігі. Оны жазушының өз ойын оқырманға оңай қабылдау мен дұрыс меңгеруді жеңілдететін әрекеті деуге болады. Мұндай бірлік ойлау нәтижесінің неше түрлі тарамдарын өзара байланыста ұстайтын тұтастық сипаттағы жағдаяттың көрінісі.

Адам баласының ойлау қабілетін тіл мен ойдың өзара байланыстылығы (қатыстылығы) тұрғысынан қарастырсақ, оны *танымдық* және *коммуникативтік* деп екіге ажыратуға болады. Ойлаудың коммуникативтік түрі деп отырғанымыз

адамның өзі үшін танылған белгілі ақпаратты басқалар үшін қайта өңдеу, яғни белгілі бір білімді тілдік қарым-қатынас ыңғайына қарай өзгерту деп түсінуге болады [Общее языкознание, 1970, 385-б.].

Мәтінді туындату кезінде ойлаудың осы аталған коммуникативтік түрі орын алатынын ескерсек, онда абзац мазмұнын ашуда да ойлаудың осы түріне сүйенуіміз керек.

Қарым-қатынас кезінде В.А.Звегинцевше айтсақ, "... мы обмениваемся не словами и их значениями, а мыслями, используя слова" [Звегинцев, 1973, 168-б.]. Яғни адамдар арасындағы тілдік қарым-қатынас қызметіндегі психикалық жағдай үзігінің санада бейнеленуінің бастамасын да "ой" деуге болады. Бірақ **ойды** нақтылау нәтижесі бір ғана сөйлем түрінде бола бермейді. Сөйлеу әрекеті бір-біріне тәуелді болып келетін бірнеше синсемантикалық сөйлемдерден тұрады. Сондықтан мәтін бойындағы "ой" жай хабарламалардың жиынтығын құрап, іріленген пішіндегі күрделі бірлік ретінде психикалық жағдай үзіктерінің қосындысының аса қомақты хабарын таратады.

Әдебиеттерде осындай бір немесе бірнеше "ойлар" қосылысының ішіндегі ең азын (минимумын) "микротақырып" ("микротема") деп атап жүр. Микротақырыптағы бірнеше ойлар денотаттың заттық қасиеттеріне (мезгілдік, функционалдық, кеңістіктік және т.б.) сай, өзара байланыста тұруы керек. Сонда ғана, барлық мәтіннің тақырыбы, аталған қатынастарда тұратын бірнеше микротақырыптардың біріккен сипаты бола алады. Әрине, ол қатынастар барлық микротақырыптарға тән денотаттардың заттық қасиеттеріне сәйкес келуі керек. Осындай жағдайда кез келген шығарманың композициялық құрылымын белгілі бір абзацтар жүйесі деп қарастыруға болады.

Әртүрлі мәтіндердің абзацтарын арнайы зерттеу нәтижесінде, олардағы сөйлемдердің бір абзац құрамына енбейтін сөйлемдермен салыстырғанда бірқатар ерекшеліктерін байқауға болады.

Олар мыналар:

а) бір абзацқа енетін сөйлемдер бір немесе бірнеше өзара байланыста тұратын микротақырыпты суреттеп тұруы қажет. Яғни абзацтың өн бойындағы сөздердің денотативті және коннотативті мағыналарының ортақ компоненттері (бөліктері) болуы керек деген сөз [Агроскина, 1972, 14-б];

ә) абзац құрамындағы сөйлемдер арнайы лексикалық құралдар арқылы өзара байланыста тұруы керек;

б) абзац құрамындағы сөйлемдердің әлеуметтік тұрғыда қалыптасқан логиканы сәулелендіретін белгілі бір синтаксистік байланысы болуы қажет.

Сөйлеу қызметіндегі адамның ойлау процесіне тән қасиет – ақиқат шындықтың құбылыстары арасында объективті түрде пайда болатын қатынастарды белгілеу. Тілде мұндай жағдай "фразадан жоғары бірлік" деңгейінде пайда болатын сөйлеудің айырықша функционалды-мағыналық типтері негізінде іске асады. Олардың ең бастыларына жататындар: суреттеу, баяндау және пайымдау типтері [Нечаева, 1974].

Сонымен, мәтін "абзацы" термині орнына "фразадан жоғары бірлік" деп аталатын мәтін бөлігін алуға болады екен.

Сөйлеу әрекетінің функционалды-мағыналық түрінің жоғарыда келтірілген үш түрін (суреттеу, баяндау және пайымдау) әрі қарай талдай түссек, мәтін абзацтарын тағы да былай қарастыруға болады:

1) хабарлау абзацтары;

2) сұрақ мәнді абзацтар;

3) ниет, ойды білдіретін абзацтар;

4) сипаттау абзацтары;

5) аралас типті абзацтар.

Құрылымдық көзқараспен қарастырғанда абзацтардың лексикалық тұрғыдан толыға түсуі, синтаксистік және грамматикалық жағынан түрлене түсуі олардың мәтіннің жалпы құрылымындағы орнына да байланысты болады. Осы себептен абзацтар: *кіріспе* (бастапқы), *ішкі* және *қорытынды абзацтар* болып саналады.

Кейбір зерттеуші ғалымдардың пайымдауынша, абзац құрылымында кездесетін "осал жерлер" де аз емес. Осыған дәлел ретінде мына жайттарды келтіруге болады:

а) мәтінді абзацтарға бөліктеуде автордың (жазушының) субъективті ұстанымы;

ә) абзацтарды тілдік бірлік ретінде санаудың қиындығы, яғни олардың тұрақты құрылымдық модельдік сипатқа ие болмауы;

б) абзацтардың бәрі бірдей аяқталған ойды білдіре бермеуі;

в) абзац көлемінің тұрақсыздығы (бір сөйлемнен бастап, бірнеше бетке дейін өсуі) [Москальская, 1981, 31-32-бб.].

Ал ғалым Чан Нгок Тхем "Абзац как грамматико-семантическая сущность" атты мақаласында абзацтың аталғандай "осал жерлерін" қабылдаудан бас тартып, оның мәртебесін "тілдік бірлік" дәрежесіне дейін жоғары көтереді. Автор абзацтарды "құрылымдық аяқталмаған", "авто-семантикалық" және "синсемантикалық" деп ажырата келіп, былай дейді: "Абзац как грамматико-семантическая сущность должен быть признан единицей языка в противопоставлении конкретному абзацу как проявлению речи" [Чан Нгок Тхем, 1977, 118-123-бб.].

Егер мәтін абзацын таңба ретінде қабылдасақ, онда сөйлеудің семиотикалық сызбасын құру үшін мына жайттарды ескеру қажет:

а) абзацтың әр тұстары әртүрлі ақпараттық "жүк" атқарады;

ә) жалпылама алғанда, абзац та ақиқат шындықтың заттық сипаты тәріздес "бірдеңені", яғни қайсыбір затты бейнелейді;

б) абзац мәтін мазмұнына белгілі бір стилистикалық және экспрессивтік элементтерді енгізеді;

г) егер қайсыбір "минимәтіннің" ("шағын мәтіннің") орнына жүретін кез келген абзац ақиқат шындықтың белгілі бір шағын оқиғасын суреттейтін болса, онда абзац денотаты $(Д_а)$ деп осы нақты жағдайдың заттар жиынтығын атайды;

д) абзац ішінде сөз болған микрожағдаят құрайтын заттар жиынтығы жайлы көрініс абзац-таңбаның десигнат $(С_а)$ компоненті болып саналады;

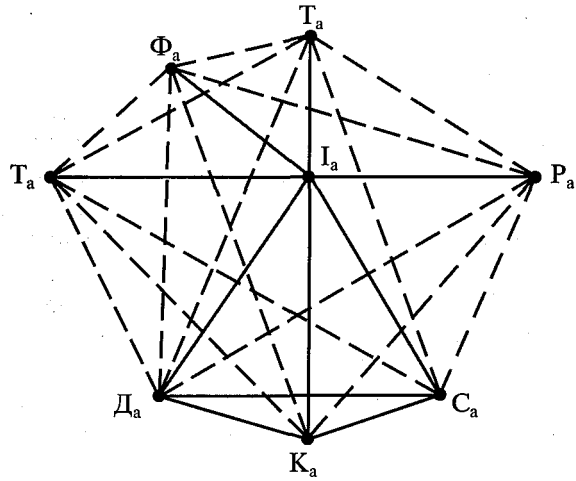
е) абзацта бейнеленетін микрожағдаяттағы заттардың қатынастық сипатын бейнелеудің әлеуметтік мәнді тәсілі ретінде – абзацтың синтаксистік мағынасы $(В_а)$ қабылданады;

ж) сөйлемнің таңбалық сипаты тәріздес, абзацтарға да ақпараттық $(Р_а)$ және коннотативтік $(К_а)$ аспектілер тән болып саналады;

з) мәтін бойындағы тақырыпшаларды біріктіріп, тұтастықта ұстайтын абзацтың тақырыптық аспекті $(Т_а)$ мен абзац аясындағы сөйлемаралық арнайы байланысты бейнелейтін фразалық аспектіні де $(Ф_а)$ ескеру қажет болады.

Сонымен, абзац-таңбаның жоғарыда келтірілген семиотикалық құрылымына сүйене отырып, абзац мазмұнына мынадай түсініктеме бергенді жөн санаймыз. Абзац мазмұны дегеніміз, ол адам санасындағы абзац атауларының сызықтық тізбегінің кішігірім жағдаяттардың санадағы осындай бейнеленуіне қатынасы.

Осы айтылғандарды сызба түрінде төмендегідей көрсетуді жөн деп таптық *(4-сурет)*.



4-сурет. Абзац-таңбаның семиотикалық құрылымы

Мұндағы:
- $Д_a$ – абзац денотаты;
- $С_a$ – десигнат;
- $К_a$ – коннотат;
- I_a – "n" тізбектен тұратын абзац атаулары;
- $В_a$ – абзацтың синтаксистік аспекті;
- $Р_a$ – ақпараттық аспект;
- $Т_a$ – тақырыптық аспект;
- $Ф_a$ – фразааралық арнайы байланыс.

1.6 Мәтін – абзац жиынтығы және таңба

Мәтіннің негізгі сипаттамасы ретінде "мазмұн тұтастығы" алынатыны ешкімге күмән туғызбайтыны белгілі. Олай болса, біз мәтін мазмұнын қалай түсінуіміз керек? Бұл жөнінен ғалымдардың көзқарасы бірізді еместігі жайлы жоғарыда да айтылды.

Сонда да зерттеуші ғалымдардың біреулері үшін мәтін мазмұны – ол "психикалық жағдай", екіншілері үшін – "тұжырымдамалық жағдай", "тұжырымдамалық сызба" немесе "тұжырымдамалық құрылым", ал үшіншілері үшін – "тақырып".

"Жасанды интеллект" мәселесімен айналысатын мамандар аталған ұғымдарды басқа атаулармен: "сценарий", "фрейм", "модель" және т.б. атап жүр.

Әрбір мәтінді, белгілі дәрежеде, жазушы санасындағы ақиқат шындықтың тұрақты бір үзігінің бейнеленуі деуге болады. Ал бұл объективті шындық мәтін авторына көптеген өзара байланыста тұратын заттардың, құбылыстардың, қасиеттердің, процестердің, қатынастардың және т.б. жиынтығы ретінде беріледі. Осындай бір-біріне тәуелді болып жатқан нысандар жиыны "жағдаят" деген ұғым арқылы түсінік тауып жүр. Құрылымдық лингвистика жайлы әдебиеттерде аталған ұғымның көптеген анықтамаларымен танысуға болады [Сильницкий, 1973, 373-392-бб.]. Біз үшін ол – белгілі түрде ұйымдасқан ақиқат немесе автор санасында өзгеріске ұшыраған қиялдау (ойлау) әлемі элементтерінің жиынтығы.

Әрине, бір ғана "ақиқат жағдай" әр адамның сана-сезімінде бірдей дәрежедегі психикалық сипатта қабылдана бермейді. Ол адам баласының "ақиқат жағдайға" байланысты ұстанымына, алға қойған мақсатына және әлемді қабылдаудағы жеке ерекшеліктеріне және т.б. жайттарға көптеп байланысты болады. Сөйтсе де, зерттеушілердің пікірінше, бір ғана ақиқат жағдайды бейнелейтін психикалық жағдаяттарға тән ортақ "бірдене" немесе "бас сипат" болады екен.

Сонымен, мәтінде белгілі бір "психикалық жағдаятты" суреттеп айту, бейнелеп түсіндіру мен жүйелеп баяндау кездеседі. Бірақ бұл жағдаяттар автордың түпкі ойына, мақсатына және коммуникациялық шартқа байланысты, белгілі түрде, өзгеріске ұшырайды; жағдаяттың көптеген құрастырушылары мен психикалық жағдаяттың қатынасына қарай автор таңдау жасап, тек қана өзінің болжамын, түйіндеу пікірін және басқа да таза логикалық ұғынысын мәтінге енгізуге ниеттенеді. Міне осындай, автор санасында өзгеріске ұшыраған "психикалық жағдаятты" оның жоғарыда аталған ерекшеліктерімен қоса, мәтіннің "мән-мағынасы", "мазмұны" немесе "тақырыбы" деп атайды.

Басқаша айтқанда, мәтіннің "мазмұны" немесе "тақырыбы" деп, жазушының санасында бейнелетін бас кейіпкер мен сипаттау нысандарының автордың ой-өрісі, мақсаты мен коммуникациялық шартына сай бейнеленуі деуге болады.

Зерттеуші ғалымдардың тұжырымдауынша, қоршаған ортаны тану ақиқат шындықты "кванттау" (бөліктеу) арқылы және олардың белгілі бір сөздер топтамасының бірлігі ретінде санамыздағы біртұтастық сипаттағы бейнеленуі деген түсінікті білдіреді. Осындай сөздер топтамасы ретінде мәтін абзацын алуға болады.

Әрбір мамандыққа қатысты ұжым мүшелерінің (әр ғылыми саланың ғалымдары, журналистер, жазушылар, ақындар және т.б.) сыртқы әлем құрылымын танудағы ортақтық сипатында абзацтарға және олардың бір-бірімен байланыста болу тәсіліне тән белгілі бір қалыптық, стеоротиптік қасиет бар.

Әрине, абзацаралық байланыстың мәтін сипатына қарай өзгеріске ұшырауы да мүмкін. Сонда да олардың толық мәтін мазмұны тақырыбына сәйкес келетін "негізгісін" тақырыптық немесе ішкі байланыс деп атауға болады [Новиков, 1983, 27-б.]. Тақырыптық қатыстылық абзац ішіндегідей, мәтін бойында да сипаттау нысанының ортақтығына қарай тұрақты сипаттағы өзекті сөздер арқылы іске асады.

Мәтін бірліктері "ішкі байланыстан" басқа қайсыбір "сыртқы байланыспен" де өзара қатынаста болады. Егер "ішкі байланысты" лексика-грамматикалық сипатқа жатқызатын болсақ, онда оны абзац ішіндегі "сөйлемаралық байланыспен" сәйкес келеді деуге болады. Абзацтарға тән "сыртқы байланыстардың" түрлері мынадай:

- эллипсистік сөйлемдер;
- бағыныңқы сөйлемдер;
- сұраулы сөйлемдер;
- есімшелі оралым арқылы жасалатын сөйлемдер және т.б. түрлері [Недбаева, 1974, 122-б.].

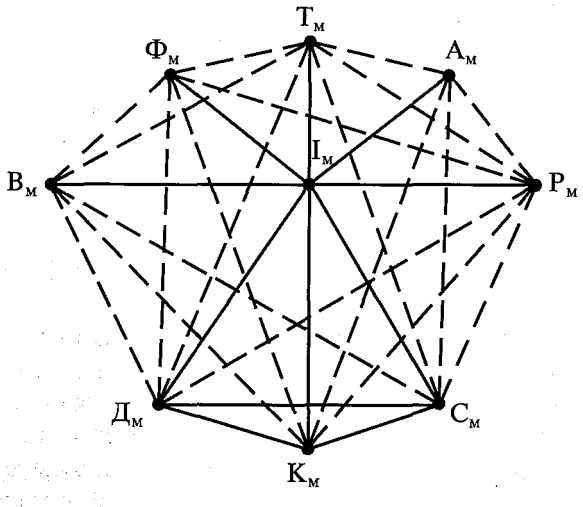
Сыртқы байланыс түріне мәтіннің жалпы синтаксистік қаңқасы ішіндегі абзац құрылымының синтаксисін тұтастықта ұстайтын "синтаксистік байланысты" да жатқызуға болады. Өкінішке орай, ондай байланыс түрі жайлы деректер өте аз. Осындай "байланыс құралы" негізінде автор мен жеке әлеуметтік топтардың ойлау логикасы арқылы бір түрлі мәтінді екінші түрлі мәтіннен ажырата алуға болады. Мысалы, Абай поэзиясы мәтінін М.Әуезов мәтінінен, математика мен физика ғылыми мәтіндерін философиялық ғылыми мәтіндерінен және т.б. ажыратуға болады.

Ақпараттық процестің жалпы құрылымында хабарламаны қабылдаушының рөлі аса маңызды деп саналады. Себебі әрбір мәтін белгілі бір мақсатпен және нақты "біреуге" бағытталып жазылады. Солай бола тұра, әрбір тыңдаушы (адресат) әртүрлі прагматикалық қасиетке ие екені белгілі. Ондай прагматикалық пресуппозициялар: тыңдаушының (адресаттың) жалпы білімін, ақпарат жайлы хабардар болуын, оның ықыласын, көзқарасы мен пікірін, психологиялық жағдаятын, мінез-құлқын, түсінушілік қабілетін адресанттың (айтушы, жазушы адамның) ескеруі. Осы айтылғандарды қорытындылайтын болсақ, ол әрбір адресант (оқырман) тиісті ақпартатты мәтін ішінен автордың ойлағанымен бірдей дәрежеде қабылдай алмайды [Бондарко, 1983, 62-б.].

Сонымен, автордың ақпаратты жолдауы және оны қабылдаушы оқырман арасындағы проблема лингвистикалық прагматика тұрғысынан қарастырылады. Ол мәтін құрудың мақсаттық ұстанымымен тығыз байланысты. Автордың тілдік қорының мәтін ішінде "ұтымды орналасуы" мәтінқұру заңдарына және мәтін синтаксисіне тікелей қатысы бар. Мәтінқұру мәселесі әртүрлі мақсатты көздеуі мүмкін: көпшілікке белгілі ақпаратты қайталап тарату, жаңа ақпаратты тарату, қабылдаушының эмоционалды қалпына әсер ету және т.б. Сөйлем мен абзацқа қарағанда, мәтіннің мұндай елеулі айырым-белгісі – оның маңызды сипаттамаларының бірі болып саналады.

Осы айтылғандарға сүйене отырып, біз, таңбаға тән қасиеттердің барлығы бірдей мәтін бойында да орын алады деп тұжырым жасауымызға болады. Бұлай қорытындылауға көптеген ғылыми әдебиеттерде келтірілген деректер дәлел бола алады [Брандес, 1979, 65-б.; Проблемы ..., 1983, 12-б.].

Сонымен, мәтін мазмұны таңба тұрғысынан мынадай түсініктеме алады. Мәтін мазмұны, біріншіден, семиотикалық құрылым бойынша мәтін аттарының (имена) сызықтық тізбегінің адам санасындағы бейнеленуі. Екіншіден, ондай бейнелеудің осы таңбаға қатысты сегіз түрлі компоненттер берілісінің санадағы бейнелену жағдаятына қатынасы. Осы тұжырым негізінде мәтін-таңбаның семиотикалық құрылымы *5-суреттегідей* көрініс тапты.



5-сурет. Мәтін-таңбаның семиотикалық сызбасы

Мұндағы:
- $Д_м$ – мәтін денотаты;
- $С_м$ – десигнат;
- $К_м$ – коннотат;
- $І_м$ – "n" тізбектен тұратын мәтін атаулары;
- $В_м$ – мәтіннің синтаксистік аспекті;
- $Р_м$ – ақпараттық аспект;
- $Т_м$ – тақырыптық аспект;
- $Ф_м$ – фразааралық байланыс аспекті;
- $А_м$ – мәтіннің мақсаттық аспекті.

1.7 Мәтін статикасы

Мәтінде көрініс тапқан кез келген жағдаят екі түрлі көзқараста қарастырылуы мүмкін. Адам баласының өмірлік және тілдік тәжірибесіне қарай жағдай элементтерін (бөліктерін) саралап алу мүмкіндігі оның ақиқат шындықты қабылдау сипатының қайсыбір ықтималдық шартына бағынуына байланысты. Мазмұн желісін өрістету үшін, автор мәтінді құрастыруға қатысты жағдаяттың ең маңызды деген басты элементтерін таңдап ала білуі қажет. Бұл элементтердің әрбіреуі денотаттардың жиыны ішінен алынатын элементтердің келесісін белгілеуге мүмкіндік туғызады [Новиков, 1983, 71-б.]. Ал мәтінде суреттелетін бас кейіпкерге немесе негізгі нысанға түрлі атаулар, мәтін мазмұнының формалды құрылымына материалдық таңбалардың сызықтық тізбегінің формалды құрылымы, яғни сөздер сәйкес келеді [Новиков, 1983, 28-б.].

Мәтін мән-мағынасының сыр-сипатын ашуға қатысты көптеген ғылыми жұмыстарды талдау арқылы мынадай тұжырым жасауға болады. Мәтіннің мағыналық тұтастығы оның өн бойында кездесетін кейбір "тірек сөздердің" болуынан. Ондай сөздер орыс тіліндегі әдебиеттерде: **опорные слова**, **топики**, **смысловые вехи**, **дискрипторы**, **ядра** және т.б. деп аталып жүр. Мұндай "тірек сөздер" құрылымында белгілі бір иерархия болатыны белгілі [Лингвистические вопросы... 1983, 123-б.]. Олардың бір бөлігі жоғарыда аталған негізгі немесе маңызды сөздер қатарына жатады және хабарламада айтылатын басты мәселені (затты не құбылысты) анықтайды. Шындығында, осы аталған **негізгі тірек сөздер** мәтінде көзделген ойдың (замысл) ықшам түрі деуге де болады. Сонымен бірге бір ғана денотаттың өзі әртүрлі сөздермен, яғни мәнмәтіндік синоним сөздермен берілуі мүмкін. Мәтін мазмұнына қатысты сөздердің сызықтық тізбек құрайтын сипатына байланысты, ондай сөздер әр абзацқа үлестіріліп беріледі.

Аталған **"негізгі тірек сөздерді"** (НТС) мәтін бойынан бөліп алуда біз мынадай формалды екі критерийді негіз етіп алдық: сөздің мәтін ішінде ең жиі қолданылуы (синонимдер мен есімдіктік айырбастарды қоса) және де сол сөздердің ең көп абзацта кездесуі.

Мұндай сөздер мәтін ішінде микрожағдаятты құрастыруға қатысы бар басқа тілдік элементтерді (етістік, сын есім, есімше және т.б.) өз айналасына шоғырландыра алатын **"орталық"** деуге болады.

Мәтінде суреттелген белгілі бір жағдаятқа қатысты бас кейіпкерлерге (суреттеу нысаны) сәйкес келетін **негізгі тірек сөздер** маңыздылық жағынан төмендеу тұратын басқа кейіпкерлерді атайтын сөздерге де қатысы болуы мүмкін. Мұндай сөздерді

екінші дәрежелі тірек сөздер немесе **қосалқы тірек сөздер** (ҚТС) деп атауды ұйғардық. Әрине, бұл ҚТС-тар НТС-тармен салыстырғанда, мәтінде аз жиілікпен және аз абзац ішінде кездеседі.

Мәтіннің НТС пен ҚТС **"Мәтіннің негізгі статикалық мазмұны"** (МНСМ) деп аталатын арнайы кесте арқылы безендіріледі.

Енді абзацтардың мән-мағыналық құрылу жағын сөз етейік. Негізінен алғанда аталған проблема лексика-семантикалық мәселеге қатысты. Себебі, абзац бірлігінің синтаксистік ерекшеліктерін айқындау оның лексикалық құрылымын жете зерттеуді қажет етеді. Тек осындай жағдайда ғана мазмұн сыры дұрыс ашылуы мүмкін [Серкова, 1980, 111-б.].

Бас тақырыптағыдай, микротақырыпшалар үшін де иерархиялық сипат пен тірек сөздер және сөз тіркестерінің қайталанып кездесу сипаты сақталады. Осымен бірге қатынастылықтың – **мағыналық субъект қатары** және **мағыналық предикат қатары** деп аталатын екі түрлі атауыштық қатары ажыратылады [Щерба, 1974, 263-б.].

Шынында да, жоғарыда айтылғандай, абзацтарда психикалық жағдаяттың қайсыбір фрагментінің бейнеленуі көрініс табады. Яғни абзацта мәтіндегі негізгі және қосалқы денотаттардың жағдаятқа қатысты маңыздылығы төмен тұратын элементтерімен байланысы және олардың қатынастық сипаты бейнеленеді деуге болады. Оларға жататындар: НТС бен ҚТС-тардың әрқилы әрекеттік және қалыптық түрлері, неше түрлі әрекет орны, әрекет уақыты, мақсат, әрекет пен қалыптың себептілігі және т.б.

Абзац сөздерінің маңыздылығына қарай сараптаудың әдіс-тәсілдері ішінен біз үшін қолайлы деп мына жағдайлар танылды. Сөздерді саралау ісінде ақиқат шындықта орын алатын жағдаяттардағы денотаттардың атқаратын мынадай рөлдері ескерілуі қажет: **"бөлік", "қасиет", "бүтін", "қабылдаушы", "әрекетші", "жиын", "құрал", "мақсат"** және т.б. [Сильдмэя, 1981, 130-б.]. Шындығында, мұндай әдіс мәтін авторының энциклопедиялық білім қорын тіркеуге қатысты іске асады. Автордың мәтін сөздігін құрастыру мәселесі мен энциклопедиялық білім қорын жинақтау проблемасы бір-бірімен тығыз қатынаста қарастырылғаны жөн.

Сонымен, абзацтардағы сөздерді саралаудағы біздің мақсатымызға сай келетін ұстаным – жоғарыда аталған "рөлдік" әдіс.

Микрожағдаят құрылымы ішінен әрі қарайғы бөлшектеуге жатпайтын құрастырушыларды да бөліп алуға мүмкіндік бар деп ұйғарым жасауға болады. Оларды **"микрожағдаяттың элементар құрастырушылары"** (МЖЭҚ) деп атайды. Әр типті мәтіндерді талдау негізінде мынадай МЖЭҚ-тарды бөліп алдық делік: **"әрекет", "қалып", "орын", "уақыт" (мезгіл), "құрал", "себеп", "мақсат", "сипаттама", "нақтылау", "тәптіштеу"** және т.б. Абзац ішінде әрбір **микрожағдаяттың элементар құрастырушыларына** (МЖЭҚ-ға) етістікке, есімшеге, зат есімге және олардың тіркестеріне қатысты **сөз** немесе **сөз тіркесі** сәйкес келеді екен. Мұндай жағдайда әрбір микрожағдаятқа тән ерекшелік белгілі МЖЭҚ тізбегі арқылы және олардың әрбіреуіне сәйкес келетін жекелік сипаттағы денотаттармен толықтыру арқылы көрініс табады. Әртүрлі мәтіндерді зерттеу негізінде мүмкін болатын абзац түрлерін жеке қарастырып, олардың әрбіреуіне, жоғарыда айтылғандай, абзацтағы микрожағдаят денотаттарының орнына жүретін "нақтылаушы-сөз" топтарын белгілі ретпен тіркеп жазу қажеттігі туындайды.

Микрожағдаяттың элементар құрастырушылары (МЖЭҚ) мәтіннің негізгі статикалық мазмұны (МНСМ) деп аталатын кестені жаңа мәліметтермен

толықтырумен қатар, белгілі дәрежеде мазмұнның қаңқасын құрайды деуге болады. Бірақ осымен мәтін мазмұны толық айқындалады деуге болмайды. Оның себебі, кез келген мәтінде тірек сөздер мен нақтылаушы-сөз топтарынан басқа тағы да "толықтырушы-сөздер" (слова-наполнители) деп аталатын сөздер тобы орын алады. Бұл сөздер осыған дейінгі мазмұндық нәтижені сапалық сипатпен толықтырады, яғни зат есім сөздермен аталатын заттың сапасын және етістік пен есімше сөздер арқылы анықталатын әрекеттің сипатын айқындай түседі. Мұндай сапалық нақтылай түсу (тәптіштеу), белгілі типтік мәтіндерде кездесу ықтималдығы мол еркін сөз тіркестерін алдын ала бөліп алу арқылы іске асады.

1.8 Мәтін динамикасы

Мәтін құрастырушылардың екінші типі ретіндегі **"мәтін динамикасы"** мәтінде сипатталған нақты жағдайда көрініс тапқан ақиқат шындықтың нысандары араларындағы қатынасты айқындайды. Бұл дәлірек айтқанда, атаулары (есімдері) мәтінде берілген денотаттарға қатысты болатын жағдайларды (процестер, әрекеттер, қалыптар, қасиеттер) белгілеу деуге болады. Адам баласының есте сақтау қабілетінің шектеулі екендігі және ол барлық жағдайларды толығымен қамти алмайтындығы белгілі. Сондықтан ол таным процесі кезінде жағдаяттардың кейбір фрагменттерін өз санасынан тыс қалдырып қояды. Ал санасында орын алған "жағдаяттарды" тұрпаттық және қатынастық ерекшеліктерімен бірге толық түрде қарастыруға мүмкіндігі бар [Чесноков, 1982, 41-42-бб.]. Әрине, болмысты фрагменттерге осылайша бөлшектеу еркін түрде жүргізілмейді, ол тек адам баласының ұзақ мезгілдік тарихи даму кезеңінде әлеуметтік тұрғыда пайдаға асқан үлгінің (модельдің) негізінде ғана іске асады [Чесноков, 1982, 42-б.].

Ойлау (мышление) мен тілдің тұтастық сипатынан санада бейнеленген болмыстың әрбір үзігіне белгілі бір тілдік тұлға сәйкес қойылады. Ондай тілдік тұлғаны кейбір зерттеуші-ғалымдар **сөз** түрінде алу керек десе, екіншілері – **сөз тіркес** немесе **сөйлем** түрінде, тіпті **толық мәтін** түрінде алғанды жөн санайды. Аталған тілдік бірліктердің ақиқат шындықтың үзіктеріне сәйкес келуін кейбір ғалымдар түгелдей жоққа шығармаса да, жекеден жалпыға қарай ғылыми қорытынды жасау мүмкіндігі мол логикалық аппаратты көбірек қолдайды.

Әр түрлі ғылыми бағыттағы зерттеушілердің байқауынша, таным процесі кезінде ақиқат шындықтың ең маңызды және жиі қайталанатын қатынастары адам санасында синтаксистік құрылым пішінінде көбірек сақталады. Мұндай құрылымдарды біздер бала шақтан бастап ана тіліміздің сөздері мен дыбыстары арқылы меңгеруге қалыптасқан болуымыз керек. Сол сияқты, сөйлеу процесі кезінде адам өз пікірін көбінесе ойланбай-ақ, дайын тілдік үлгілерге еліктеп айтады [Адмони, 1976, 3-б.].

Сонымен, мұндай үлгілер (шаблондар) басқа тілдік деңгей бірліктерімен бірге, мәтінді құрастыру процесіне қатынасатын, біздің санамызда орын алатын психикалық ақиқат. Орыс тіліндегі ғылыми әдебиеттерде аталған үлгілер мынандай терминдермен: "синтаксические модели", "структурные схемы предложений", "конструкции", "фразовые стереотипы", "синтаксические формы" және т.б. аталып жүр.

Қазіргі психология ғылымының деректері бойынша ойды вербальды (сөзбен) нақтылау нәтижесі тек сөйлем күйінде болуы міндетті емес. Сөйлеу әрекеті өзара тәуелділікте тұратын бірнеше сөйлемнен де тұруы мүмкін [Зарубина, 1968, 146-б.].

Яғни мұны, табиғи тілімізде, ой құрылымының сөйлемнен тыс жатқан қатынасын білдіретін белгілі бір мағыналы-құрылымдық үлгісі болады деп ұғыну керек сияқты. Әдетте, мұндай өзара тәуелді ой тізбегі *айтылым* (высказывание) деп аталып жүр.

Сонымен, айтылым ой-дискурстердің бір-біріне ұласу заңдылығымен жасалады деуге болады. Мұнда негізгі ереже ретінде пайымдаудың логикалық жалғастығы, дәлелділігі, түйінділігі немесе қисынды суреттілігі алынады [Колшанский, 1980, 32-б.].

Айтылымдағы бірліктердің негізгі ерекшелігі бөліктен бүтінді құруда олардың анық және белгілі дәрежеде тұрақты тұрпатқа ие болатындығы. Теориялық жағынан алғанда, мұндай бірліктерді қолдану жағдайында, олардың шығу табиғатын біле бермеуіміз де мүмкін. Себебі, тілді үйрену мен өмірге бейімделу кезінде, тілдік бірліктер болмыс үзіктерімен сәйкес келуімен қатар, автордың жеке басының тәжірибесін және бүкіл қоғамның әлеуметтік тәжірибесін де қоса алып жүреді.

Аталған мәселені терең түрде зерттеу нәтижесінде бірнеше сөйлемді қамтитын синтаксистік үлгілердің (шаблондардың) болуы болмыстың санада бейнеленуінен озып шығу ілімімен де түсіндіріледі [Анохин, 1962, 14-25-бб.]. Бұл теория бойынша — "человек обладает способностью на основе врожденного и приобретенного опыта предвидеть развитие ситуативных обстоятельств и настраиваться на определенную реакцию с некоторым упреждением вероятностного хода событий" [Колшанский, 1983, 50-б.].

Сөйлеу тәжірибесіндегі айтылым-шаблондардың тиісті түрде орын алуын тілшілердің, психологтардың, психолингвистер мен тіл үйрету мамандарының да зерттеулерінен кездестіруге болады. Сонымен бірге, мұндай "шаблондардың" жиынтығы шектеулі және жеке автор мәтіні мен шағын пәндік ая үшін сан жағынан аз, әрі ерекше сипатта болады [Солнцев, 1977, 297-б.].

Жоғарыда сөз болғандай, мұндай бірнеше сөйлемдерді біріктіріп тұрған синтаксистік тұтастық "фразадан тыс бірлік" немесе "күрделі синтаксистік тұтастық" және т.б деп аталып жүр.

Соңғы жылдардағы зерттеулерге сүйенсек, мәтін ішіндегі абзацты да ақиқат болмыстың **"логикалық көшірме бейнесі"** деуге негіз бар [Серкова, 1980, 111-б.]. Бұл айтылғандарды тек лингвистердің зерттеу нәтижелері ғана растап отырған жоқ, сонымен бірге психологтар мен психолингвистердің де тәжірибелері дәлелдеп отыр. Мысалы, бірінші сынаққа қатысушы адамдарға бір ғана мәтін әртүрлі жолмен дайындалып ұсынылған. Олар мыналар: автордың абзацқа бөлуі қалпындағы мәтін; абзацқа бөлінбеген пішіндегі мәтін; логикалық құрылымы ескерілмей, абзацтарға кездейсоқтық сипатта бөлінген мәтін. Тәжірибе нәтижесінде сынаққа қатысушылар бірінші мәтін (авторлық қалып) мазмұнын өте сенімді әрі жеңіл және тез түрде қабылдаған. Ал тәжірибенің 2-ші және 3-ші жолдарында сынаққа қатысушылар мәтін мазмұнын аса көп уақыт жұмсап, нашар қабылдаған, ал кейбіреулері мәтін мазмұнын тіпті дұрыс түсіне де алмаған.

Екінші сынаққа қатысушы адамдарға үш түрлі стильде жазылған мәтіндер ұсынылады: көркем әдебиет стилі (көктемді суреттеу), ғылыми стиль ("қасқырды" ғылыми тұрғыда суреттеу) және ғылыми көпшілікке арналған стиль ("темірге" қатысты тақырып). Аталған мәтіндер абзацтарға бөлінбеген қалпында, тек сөйлем тізбектері пішінінде ұсынылған. Тәжірибеге қатысушылардың, өз қалауынша, осы мәтіндерді абзацтарға бөлшектеумен қатар, олардан қиысқан мәтін құру талабы қойылған [Бондаренко, Шрейдер, 1978, 80-91-бб.].

Тәжірибенің нәтижесінде: көркем әдебиет стилі бойынша сынаққа қатысушылардың барлығы бірдей дерлік сөйлем тізбектерін ұқсас мағыналық топтарға біріктірген. Ал олардың мәтінді абзацтарға бөліктеу арқылы қиысқан мәтін құруларындағы айырмашылық, негізінен, тек сөйлем тәртібін реттеу жолдарында ғана байқалған. Сынақ жүргізу нәтижесінде қатысушылардың үштен бірі ғана абзацтарға мағыналық топтауда сөйлемдер ретін мәтін авторымен сәйкес жүргізген. Ал басқа жағдайларда мағыналық топтардың орындары ауысып, авторлық мәтін айтарлықтай өзгеріске ұшыраған.

Екінші стиль бойынша, яғни ғылыми суреттеу мәтінінде сынаққа қатысушылар деректерді қатаң логикалық тәртіппен орналастырып, барлығын да түпнұсқаға ұқсас топтастырған [Бондаренко, Шрейдер, 1978, 82-83-бб.].

Үшінші, ғылыми-көпшілікке арналған мәтін, жоғарыда аталған мәтіндерден пайымдауға қажетті заттық нысандар санының көптігімен ерекшеленеді. Сондықтан қатысушылардың әрбіреуі бір сөйлемнің өзін әртүрлі мағыналық топтарға жатқызғаны байқалады. Дегенмен, мәтіндегі ірі мағыналық топтар абзацтарға дұрыс топталып, автор ұсынған мазмұн айтарлықтай өзгеріске ұшырамаған деуге болады [Бондаренко, Шрейдер, 1978, 84-б.].

Сонымен, осы және басқа да тәжірибелердің нәтижелеріне сүйенетін болсақ, автор (қалам иесі) мәтінді абзацтарға бөлуді қалай болса солай жүргізбейтіндігіне көз жеткізуге болады. Ол абзацтар белгілі ережемен және ғылыми мәтіндерде әрі қатаң, әрі стереотипті дәрежеде, ал көркем әдебиет мәтіндерінде барынша еркін және шығармашылық сипатта болатынын атауға болады. Осымен бірге, сөйлемдерді абзацқа жинақтаудың әр авторға тән ерекшеліктері де жоқ емес. Мұндай ерекшеліктер жеке шағын пәндік салаларға қатысты мәтіндерде де орын алады [Перекальская, 1980].

1.9 Мәтінді құрастырудағы ықтималдық факторлардың рөлі

Жоғарыда қарастырылған мәтін және онда бейнеленген болмыс жағдайының қарым-қатынасы мәтін мазмұнының құрылу жағдаятын толық түсіндіре алмайды. Мәселен, бір ғана нәрсе жайлы әр жолмен, әртүрлі айтуға болатыны бізге белгілі. Бұған себеп болатын айтушының (жазушының) пиғылы, автордың мәтінді құру мақсаты, мәтін типі, автордың әлемді қабылдаудағы субъективтік ерекшелігі, көптеген басқа да факторларға қатысты автордың әлеуметтік және тілдік білім дәрежесі және т.б. екендігіне ешкім күмән туғызбас. Осындай факторлардың атқаратын рөлі мен өзара қатынастылығын айқындау тіл қызметіне жаңаша тұрғыда, яғни қажеттілік пен кездейсоқтық, себептілік пен ықтималдық көзқарасымен қарауға мүмкіндік тудырады.

Табиғи тілдің әлеуметтік және ұдайы даму сипатындағы жүйе болуынан онда белгілі бір ықтималдық қатынастар орын алуы қажет. Бірақ тілдің өз иесінсіз өмір сүрмейтінін ескерсек, онда ол әлеуметтік қоғам мүшесінің әрбіреуінің "басында бар", әр қоғам мүшесіне ортақ құбылыс. Сол сияқты, әрбір қоғам заманның өзгермелілігіне тәуелді. Сондықтан өзгермелі замандағы оқиғалар да құбылмалы сипатта болып, өзгеріп тұрады, бірақ олардың өмірде пайда болуы не пайда болмауы белгілі бір заңдылыққа бағынады. Оған жыл мезгілінің ауысуын, күн мен түннің ұзаруын, ауа райының құбылуын және т.б. мысалдарды көптеп келтіруге болар

еді. Ал көптеген оқиғалардың өмірде кездесуі не кездеспеуі, пайда болуы немесе пайда болмауы қайсыбір ықтималдық дәрежесіне сәйкес келеді. Сол себепті, әрбір қоғам мүшесі өзі өмір сүретін ортаны белгілі бір ықтималдықпен қабылдай алады және ол мәліметтерді өзінің мінез-құлқы мен жүріс-тұрысын реттеуге пайдалана алады [Прогноз ..., 1974, 8-б.]. Адам санасында өткен өмір тәжірибесінің бейнеленуі оның есінде әртүрлі (өзі қатысқан) оқиғалардың кездесу жиілігіне қатысты мәліметтердің жиыны негізінде іске асады.

Адамның өмір тану процесінде оның нерв жүйесіне бір мезгілде бірнеше түрлі нерв импульстері әсер етеді. Олар қоршаған орта факторларының әсері жайлы адам ағзасына сигнал береді. Сондықтан адам ағзасының орта әсерінің түрлеріне қарай, оған қалай жауап беретіні жайлы мәліметті толық сеніммен айту оңайға түспейді. Мұндай жауапты (реакцияны) қоғам мүшесінің жекелік сипатына қарай, болжам ретінде, белгілі бір ықтималдықпен ғана сөз етуге болады. Сонымен қатар, адам баласының сөйлеу қызметі де, психикалық қызметтің кез келген пішіні тәрізді, негізінде, психологиялық құрылымға ие болатындығы дәлелденіп отыр [Леонтьев, 1977, 101-б.]. Сондықтан адамның сөйлеу әрекетінде оның мінез-құлқындағыдай ықтималдық құрастырушыларды кездестіруге болады. Яғни бұл, жеке бастың сөйлеу тәжірибесі де ықтималдық құрылымға ие деген сөз [Прогноз ..., 1974]. Тағы да бір ескеруге тұратын жайт, ол адамның "өзіне-өзі сөйлемейтіндігі", яғни әр уақытта сөйлеудің біреуге бағытталуы және оның белгілі бір өмірлік жағдаятқа қатысты болуы. Көптеген зерттеу нәтижелері бойынша, айтушы жағдаят моделін қабылдауда оны эталон ретінде қалыптастырады. Адамның қабылдау мүмкіндігі, біріншіден, оның бұрынғы басынан кешкен әлеуметтік тәжірибесіне байланысты болса, екіншіден, атқаратын қызметіне қатысты шешетін мәселелерімен және олардың себеп-салдарымен (мотивімен) байланысты болады. Басқаша айтқанда, жеке адам психикасының әлеуметтік сипатта болуынан тілді әлеуметтік тәжірибені меңгеру құралы деуге болады [Сорокин, Тарасов, 1979, 135-б.].

Сонымен, қазіргі лингвистикада тіл мен диалект арасындағы парадокс (оғаш пікір) былайша шешім табуда: "Все сугубо индивидуальное ... при функционировании индивидуального языка, входя в речь, не перестает быть языком – пока оно социально, т.е. принадлежит и другому и лишь поэтому понимается. Именно в этом смысле все в речи индивида и тем самым весь индивидуальный язык детерминирован социально" [Леонтьев, 1964, 162-б.].

Әр бағыттағы мамандардың зерттеулері бойынша, тіл қызметі процесінде ықтималды-статистикалық заңдылық орын алады. Ал бұл жағдай (факт) зерттеушілерден көптеген тың мәселелерді шешуді қажет етеді. Мысалы, оларға жататындар: заңдылық қандай тілдік деңгейлерде орын алады; заңдылықтың орын алу тетігі неде; себепті (детерминді) және ықтималды процестердің қажетті арақатынасы қалайша қамтылады және т.б.

Әрине, бұл сұрақтарға жауап беру үшін әр типтегі нақты мәтіндер бойынша ықтималды-статистикалық талдаудың ғана негізінде жауап беруге болады.

Аталған мәселе "ықтималдық программалау" арқылы орындалады. Адам санасына жаңадан "түсіп жатқан" тың ақпаратты, оның өткен өмірлік тәжірибесі бойынша есінде бұрыннан жинақталған ақпараттармен салыстыруға мүмкіндігі бар. Осындай салыстырма деректер болашақ жағдайларды болжауға және оларға қайсыбір ықтималдықтарды сәйкес қоюға мүмкіндік тудырады [Файгенберг, Иванников, 1978, 81-б]. Мұндай жалпыфизиологиялық заңдылық сөйлеу қызметіне де таралады. Яғни адам баласы сөйлеу элементтерінің қайсы бірінің қолдану

ықтималдығы туралы жеке өзі хабардар бола отырып, осының негізінде, алдында тұрған сөйлеу жағдаятының ықтималдық құрылымының субъективті моделін құра алады. Осы модельдің негізінде жеке өзінің болжам құруына және оған сәйкестендіріп, сөйлеу әрекетін ұйымдастыруға мол мүмкіндік туындайды [Прогноз ..., 22-б.]. Ал адам психикасы өзінің табиғатында әлеуметтік болуынан әрбір индивидтің тілдік қабілеті де әлеуметтік сипатта болады. Бұған Г.О.Винокурдың: "В каждом тексте может быть выделено нечто внеличное, общее, надиндивидуальное" деп пайымдауы дәлел бола алады [Винокур, 1959]. Аталған әлеуметтік сипат іс жүзінде мына жағдайдан байқалады: адам баласы мәтінді туындату процесінде сөздерді, сөз тіркестерін, сөйлемдерді, тіпті олардан да ірі бірліктерді шексіз түрде емес, белгілі бір шекаралық арада қолданады. Мұндай "шекара" әрбір жеке адамға тән болуымен қатар, оның барлық өмір сүру мезгіліндегі көптеген басынан кешкен факторларға байланысты. Тілдік қатынасқа қажетті мұндай бірліктер жиынының өзегі адамның бала шағынан қалыптасады екен. Яғни сөйлеу тәжірибесінде баланың үлкенге еліктеуі және бала тілінің қалыптасуындағы үлкендердің белсенді түрде араласу әсерлері деуге болады [Разинкина, 1978]. Сөйлеу қызметіндегі стереотиптік немесе қалыптық сипат автордың да және қабылдаушының да ми қызметіне салмақ түсірмеуге және оның күшін үнемдеуге бағышталған. Стереотиптікте құбылмалылық сипатының аз болуы тілдік құралды еркін таңдау мүмкіндігін азайтады да оны автоматты түрде таңдауға мүмкіндік жасайды. Фонематикалық, ауызша фразалық және мәтіндік стереотиптер түрлері ажыратылады [Проблемы... 1983, 90-93-бб.]. Бұл стеротип түрлеріне тән ерекшелік – әр зерттеу деңгейіне байланысты олардың "қаталдық" сипатының бәсеңдеуінен мақсатқа жетудің мүмкіндігінің артуы.

Ауызша стереотиптің негізінде мынадай жайт жатқаны аян: шағын пәндік сала мен бір жағдаятқа қатысты тек бір ғана ұғымның түсінігін беру үшін бірдей мүмкіндіктер (құралдар) пайдаланылады. Ал фразалық стереотип дегеніміз, ой дамуының қайсыбір жалпы сызбасын адам есінен "шақырып алу", яғни оның бұрын қалыптасқан "сөйлеу өрнегін" қайтадан пайдалану деген сөз [Колшанский, 1983, 46-б.].

Фразалық стереотип негізінде әрбір айтылымға қатысты қайсыбір лексикалық элементтерді маңызды деп санау (актуализация), одан басқаларын автоматты түрде, ешбір күмәнсыз таңдауға мүмкіндік туғызады. Жоғарыда абзацтың маңыздылығы мен оның мәтін құрылымындағы рөлі және ерекшелігі жайлы келтірілген пайымдаулар "фразадан жоғары" стереотиптіктің болатындығын да дәлелдей түседі.

Енді мәтін стереотипіне келсек, ол туралы "Проблемы текстуальной лингвистики" атты ғылыми еңбекте былай түсіндірілген: "Особый самоорганизующийся психологический механизм, способствующий "укладыванию" поступающей из текста разнообразной информации в русло разветвленной структурно-смысловой схемы в условиях взаимодействия ретроспекции и антиципации" [Проблемы ..., 93-б.].

Ал мәтінді туындату (порождения текста) мәселесіне қатысты мәтіндік стереотип ұғымын өзін-өзі туындата алатын айрықша психологиялық тетік деп түсіндіруге болады. Аталған тетік бойынша көпмағыналы құрылымдық-мәнді сызбаны және мәтін өрнегін (формуласын) құруға болады. Бір ғана мазмұнды баяндауға қатысты мәтін өрнектерінің саны мәтін типіне және автордың тілдік білім деңгейіне қатысты болып келеді. Ал бұл жердегі мәтін иесінің тілдік білім деңгейі жеке тұлғаның әлеуметтік іс-тәжірибесіне сәйкес келетіні белгілі.

Әр ғылыми бағыттағы мамандардың байқауынша, мәтіннің туындауы және оның мазмұнын түсіну жеке тілдік элементтер бойынша емес, белгілі бір "блоктар" (топтамалар), "кванттар" арқылы жүргізіледі. Мәтін ішіндегі осы тәріздес белгілі семантикалық арақашықтықта тұратын алдыңғы бөліктер кейінгілерімен өзара байланыста болады [Жинкин, 1978, 51-б.].

Сөйлеу қызметінің, ең алдымен, заттық мазмұннан тұруына байланысты тілдік қарым-қатынас та жіктелген деуге болады. Мәтінді қайсыбір "кесектерге" бөліктеу ақиқат шындықты "логикалық көшірмеге" бөліктеуге негізделеді. Ал ол адам баласының жалпылама ғылыми таным-тәжірибесі арқылы іске асады [Серков, 1980, 111-б.].

Сол сияқты, адам қызметінің сыртқы және ішкі тарамдарының жалпы құрылымының бірдейлігінен, жоғарыда сөз болған мәтінді құрастырушыларға бөлшектеу тәсілі биология мен психология салаларының деректерімен де үйлесім табады [Леонтьев, 1977, 101-б]. Мысалы, жануарлардың мінез-құлқы мынандай элементар әрекеттердің жиынтығынан тұрады: "қашу", "құдалау", "індету", "тамақ іздеу", "көңілін табу", "жағымпаздану" және т.б. Сонымен бірге, әрбір жануарлар типінің өзіне ғана тән тұрақтанған қимыл-әрекеттері бар. Бұл әрекеттерді бір-бірімен қиыстыру арқылы жануарлар күрделі түрдегі мінез-құлықтарын білдіре алады [Бернштейн, 1966, 79-б]. Дәл осы айтылғандай, ой әрекетінің күрделі тұрпаты да элементарлық ақпараттар процестерінің жиынтығынан тұратыны анықталып отыр. Элементар ақпараттарды қиыстыру арқылы ой әрекетінің де әртүрлі тұрпаттарын туындатуға болады. Мысалы, шахмат ойыны, есеп шығару, бейнетану, бір тілдегі мәтінді екінші тілге аудару және т.б.

Осындай әрбір "кірпіштерді" тұтастық қалыпқа келтіру мәселесі мынандай үш принцип негізінде іске асады:

1) ықтималдық принцип, яғни бөлек элементтер кездейсоқ біріктіріледі;

2) себептілік (детерминдік) принцип – біріктіру кезінде "кірпіштер" реті бірізді және қатаң тәртіппен беріледі;

3) ықтималды-алгоритмдік, яғни "кірпіштерді" біртұтас етіп біріктіру кезінде байланыстың кездейсоқтық және детерминдік (алгоритмдік) сипаттары ескеріледі.

Біріктірудің соңғы принципі, яғни ықтималды-алгоритмдік, ақиқат әлемге тән негізгі принциптер қатарына жатады [Жуков, 1973, 68-б.].

Бұрын да айтқанымыздай, ақиқат әлемінің күрделі құрылымдарында ықтималдық құрастырушылардың болуы мұндай жүйелердің өзгермелілігі мен жаңжақты қасиетін білдіреді.

Мәтінді топтамалардан (блоктардан) құру мен мазмұнды "кванттау" проблемасына қатысты айтатынымыз, ондай бөлшектеудің мәтін типіне (ғылыми, көркем әдебиет және т.б.) тәуелді болуы, бағышталуы, автордың ниетіне қатыстылығы, автордың тілдік білім дәрежесіне және басқа да көптеген факторларға байланысты болуы. Мәтіннің семантикалық құрылымын зерттеу барысында мәтіннен әртүрлі мағыналық "кесектерді" бөліп алу процесі жүргізілген болатын. Ол "кесектер" орыс тіліндегі әдебиеттерде **"текстовые элементы"** және **"блок-мотивы"** [Школьник, 1978, 209-б], **"фрагменты"** [Новиков, ... 1983, 200-б.], **"аспекты"** [Гендина, 1979, 7-б]. **"сгустки"** [Скороходько, 1983, 138-б], **"абзацы"** және т.б. деп аталып жүр. Жоғарыда сөз болғандай, **"абзац"** мәтіннің мағыналық **"кванты"** ретінде және **"қатып қалған" синтаксистік үлгі** (шаблон) ретінде **ең кіші** (минимум) **логика-семантикалық компонент** саналады. Мәтінді абзацтар арқылы құрастыру мәтінді жеңіл қабылдауға, автордың мәтінді құрудағы ниеті мен мақсатына сай, бір ғана мазмұнды әр вариант түрінде оқырманға ұсынуға мүмкіндік береді.

Абзацтардан құралған біртұтас мәтінді қалыптастыруда ықтималдықты қалай қолдану жолдары әзірше анық емес. Мысалы, ғылыми стильдегі мәтіндер үшін, олардағы логикалық құрастырушылардың артықтығынан мәтін абзацтары стереотипті тұрғыда сан жағынан шектеулі сызба бойынша "сапқа тұра қалады" деуге болады. Ал көркем әдебиет мәтіндеріндегі абзацтар еркін түрде орналасу сипатымен ерекшеленеді [Серкова, 1981, 120-б]. Сонымен бірге, авторға тән абзацтар құрылымы мен олардың өзара орналасу тәртібі болады деген жорамалдар да айтылып жүр [Невежина,... 1971; Сильман, 1975, 216-б.].

Әрине, әрбір мәтінде әлеуметтік-лингвистикалық себеппен болатын стереотиптік элементтерден басқа, жеке автор тіліне тән ерекше қолданысты да кездестіруге болады. Олар жазушылар мен ақындар үшін ойлаудың жекелік сипатындағы әсерлі және эстетикалық мазмұн құру құралы, ал ғалымдар ондай элементтерді мәтіннің байланыстылығы мен қисындылығы үшін қолданады. Мысалы, ғылыми мәтінде ондай жекелік авторлық "із" сол ғылым саласы атауларын шеберлікпен ұсына білу әдісі. Солай бола тұра, сөйлеудегі жекелік сипат, сол стиль нормасынан шет шыға алмайды және әрбір мәтін тілдік жүйеге заңдылығына бағынышты болады.

Қорыта айтқанда, тілдік жүйеге сәйкес келетін жағдаяттар: "қайталанған" мен "жаңадан жасалған" және "қайталанатын" мен "жаңадан жасалатын", яғни барлық осы мәтінде бар және одан тыс жатқан тілдік жүйе элементтері.

Екінші тарау
ТІЛДІ МОДЕЛЬДЕУ

2.1 Табиғи тіл – әлем моделі

Тілді модельдеудің нақты проблемаларын қарастырудан бұрын, мына үш жағдайға көңіл бөлейік [Откупщикова, 1996, 100-112-бб.]:

1. Адамның сөйлеу қызметін үш түрлі тип ретінде қарастыруға болады:

1-ші тип – адам өмірі мен оның қызметіндегі орын алатын әдеттегі тілдік қатынас (коммуникация);

2-ші тип – бір тілден екінші тілге аудару кезіндегі тілдік қызмет;

3-ші тип – тілді үйрету немесе әр ғылым саласына (мәдениетке, өнерге және т.б.) қатысты білімдерді үйрету (үйрену) кезіндегі тілдік қызмет.

2. Сөйлеу қызметінің барлық типтерінде адамдар бір-бірімен қарым-қатынас (коммуникация) кезіндегі ауызша және жазбаша тілдік шығармаларды (немесе мәтіндерді) түсіну мен нәтиже шығару (продуцировать) ісімен айналысады.

3. Мәтіндерді түсіну мен одан нәтиже шығару нақты тілдің тілдік тетігін білуге негізделеді. Яғни тілдің лексикалық құрамын, сөзжасам мен сөз өзгерту ережелерін, сөйлем мен толық мәтіндерді құрастыру ережелерін білуді қажет етеді.

Осылайша тіл қызметін үш түрлі қарастыру негізінде тілді модельдеуді де үш түрге бөліп қарастыруға болады.

Модельдердің бірінші түріне – тіл қызметінің түріне қарай "қайта жасау" (воспроизводящие) модельдері жатады. Бұлар: мәтін мазмұнын түсіну немесе "талдау" (анализдеу) модельдері, тілдік шығармаларды туындату немесе "жинақтау" (синтездеу) модельдері, "автоматты аудару" модельдері, "тілге үйрету" модельдері, табиғи тілде диалог түрін пайдаланатын – "әр типтік ой өрісі жүйелерінің" (интеллектуальные системы) модельдері және т.б.

Модельдердің екінші түрі – тілдік шығармалар (мәтіндер) модельдері деп аталады. Модельдеу проблемасының аспектісіне қарай олардың ерекшеліктері мынада: тілдік шығармалар, яғни мәтіндер тіл қызметінің кез келген типінің міндетті түрдегі компоненті болуынан, кез келген модельдің де міндетті құрамы болады. Солай бола тұра, тілдік шығармалар (немесе мәтін) болмыстың белгілі бір моделі немесе "әлем моделі" бола алады.

Үшінші модельдер түрін – тілдік шығармаларды түсіну мен оларды құрастыруға қажетті тілдік заңдылықтарды айқындайтын "тілдік механизмдер моделі" деп атайды.

"Табиғи тіл – бізді қоршаған шындықтың моделі" деген тұжырымды қазіргі кездегі тілшілердің көпшілігі қолдайды. Дегенмен, мұндай тұжырыммен тек тілдік шығармалар (немесе мәтін) жайлы сөз болған жағдайда ғана келісуге болады.

Әлемді модельдеуде тілдік тетіктің (немесе тілдік жүйенің) де қатынасы бар, бірақ ол жаңаша түрде сипатталады.

Тілдік шығармалардың әлемді модельдеуге қатыстылығын, модельдердің жалпы қасиеттерін талдау мен олардың тілдік шығармалар талабына сай келуін анықтау негізінде көз жеткізуге болады.

Жалпы жағдайда, модельді ақиқат шындықтың қайсыбір жүйесін белгілі бір тәсілмен қайта өндірілетін конструкция (құрылыс) деп түсіндіруге болады.

Енді модельдердің жалпы қасиеттерін және оларға елеулі түрде әсер ететін модельдеу жағдаятты атаумен қатар, оларды әлем моделі ретіндегі тіл қасиеттерімен және осы жағдайға тән келетін модельдеу жағдаятымен сәйкестендіре қарастырайық.

2.1.1 Жағдаятты модельдеудегі дуалистік сипат

Жағдаятты модельдеудің кез келген әдісі, ең алдымен, модельдеу нысаны мен оны өмірде ақиқат түрде кездеседі деп болжам жасауға қатысты болады. Табиғи тіл тұрғысынан алғанда, модельденетін нысан ретінде адам баласын қоршаған орта, яғни ақиқат шындық, ал модель ретінде оны суреттейтін мәтіннің өзі алынады. Бұл жағдайда әлемнің моделі ретіндегі тілдің ерекшелігі – ақиқат шындықтың кең мағынада түсінік табуы. Яғни ондай түсінік бойынша: әлем тек физикалық сипатқа ғана ие емес, ол адам өмірінің интеллектуалды (ойлау), рухтық, әлеуметтік және адамның өзі туындатқан тілдік шығармаларды қоса есептегендегі барлық құндылықтарға ие болуында.

Тілдің әлем моделі ретіндегі тағы бір ерекшелік сипаты, оның ақиқат шындықтан тыс жатқан: мүмкіндік, ойлау, армандау, қиялдау, ықтималдық және жалған әлемді мәтін арқылы модельдеудің шексіз мүмкіндігінің болуында.

Француздың белгілі лингвист-ғалымы Ш.Балли жазғандай, "Действительность может быть не только объективной но и мыслимой, воображаемой" [Балли, 1955, 88-б]. Мұндай мүмкіндіктер ертегілерде, аңыздарда, мифтерде, фантастикалық, тіпті, шындықты суреттейтін көркем әдебиет шығармаларында да орын алады.

2.1.2 Модельдеу нысаны және модель

Модельдеуге қатысты жағдаяттың жүйелілігі. Модель ұғымының анықтамасы бойынша, модель қайсыбір жүйені қайта туындатады. Ондай жүйе өзара байланыста тұратын және бүтіндік сипатта қызмет ететін элементтер немесе бөліктерден тұрады.

Тілдік шығармаларда, яғни мәтіндерде ақиқат шындықтан тұратын жүйені тілдік тұрпатта қайта өндіреді. Мұндағы ерекшелік – үлкен жүйелерді ғана суреттеу емес, мысалы, космос әлемін, сонымен бірге кішігірім жүйелерді де суреттеу мүмкіндікке ие болуы. Кішігірім жүйенің мысалы ретінде қысқаша диалог түріндегі тілдесуді суреттейтін күнделікті адам өмірінің микроәлемін айтуға болады.

Кез келген модель бір-бірімен белгілі қатынастық байланыста тұратын элементтер жүйесін құрайды. Әрбір мәтін де өзара мағыналық, құрылымдық және сызықтық қатынастардың байланысын құрайтын элементтер жүйесінен тұрады.

Модельденуші нысанның модельге қайта туындату аспектісі. Қайта өндіру (воспроизводство) аспектісі модельдеудің міндегі мен мақсатына байланысты болып келеді. Модельденуші нысанның қандай аспектісі модельге қайта тундатылуына қарай, модельдер: *құрылымдық, функционалды және аралас модельдер* деп ажыратылады.

Құрылымдық модельдерде модельденуші нысанның құрылымдық жағы туындатылады, яғни оларға жататындар – жүйе ішіндегі маңызды элементтер мен өзара қатынастылық.

Функционалды модельдерде нысандардың табиғаты мен қоғамдағы сипаты қайта туындайды. Ал аралас модельдерде аталған аспектілер біріккен сипатта қайта туындайды.

Тілдесу (сөйлесу) түріндегі шығармалар (речевые произведения) құрылымдық модельдердің айырықша түріне жатады. Егер тіпті, тілдесу (сөйлесу) түріндегі шығарма, болмыстың қайсыбір кішкене бөлігінің ғана жүйелік қызметін суреттеп, құрылымдық модельдің қызметін атқарса да, ол функционалды модель бола алмайды.

Модельдер түрлері, жалпы алғанда, *интеллектуалды* және *физикалық* деп екіге ажыратылады. Ал тілдесу түріндегі шығармалар моделі интеллектуалды-құрылымдық модель түріне жатады.

Модельдердің дуалистік (екі жақты) сипаты. Кез келген модель қайсыбір тұрпат пен мазмұнның бірлігі ретінде көрініс табуынан ол дуалистік сипатқа ие болады деп саналады. Бұл жердегі айта кететін жайт, модельдің тұрпаты мен мазмұны модельденуші нысанның тұрпаты мен мазмұнынан айрықша болады. Мұндай жағдай, тіпті күшті еліктеушілік сипаты негізінде болатын қайта туындату процестерінде де айырмашылықтары бар деп саналады. Мысалы, су кемесінен айнытпай жасалған модельдің тұрпаты түпкі нысаннан (нағыз кеменің өзінен) өлшемдік жағынан, құрамдық бөліктерді нақтылау дәрежесінен, жасалған материалынан және т.б. айырмашылық табады.

Интеллектуалды (ой әрекеті) модельдердің тұрпатын табиғи және жасанды тілдер негізінде жасалған суреттеу көрінісі десек, ал оның мазмұнын элементтер жүйесі мен олардың арасындағы қатынастар деп ұйғаруға болады.

Әлем моделін сипаттайтын мәтіннің тұрпаты деп, уақытқа қатысты (сөйлеу тілі үшін) немесе сызықтық (жазба тілі үшін) пішініне қатысты мәтін элементтерінің тізбегін атайды. Ал оның мазмұны ретінде – мәтіннің грамматикалық және мағыналық құрылымы алынады. Тілдік шығармалардың екі жақтық сипаты тілдің таңбалық табиғатымен түсіндіріледі.

Модельдегі модельденуші нысанның жинақты қорытындысы және оның жеңілдетілген тұрпаты (пішіні). Модель әр уақытта модельденуші құбылыстың құрылымдық, функционалдық немесе басқа да бір қасиеттерінің қайсыбір тәсілдермен жинақты қорытындыланған және сонымен бірге, жеңілдетілген пішіні. Сондықтан модель ұғымын нысанның немесе олардың кластарының жинақты қорытындыланған сипаты деп ұғынған жөн. Әрине, модельденуші нысанның кейбір өлшемдері мұндай модельде анықталмай қалуы да мүмкін.

Коммуникация мақсатына сай, табиғи тіл негізінде **әлемді модельдеу** әртүрлі дәлдікпен орындалады. Бірақ қалай болғанда да, **әлем моделі** аса бір қарапайым және жинақты қорытындыланған күйінде көрініс табады. Тілшілер мен логика мамандарының тілдік айтылым мағыналарының шамамен алғандағы немесе белгісіздікке қатысты пікірлері көпшілікке белгілі. Мысалы, көптік жалғаудың белгісіздік мағынасы: *Аулада кішкентай балалар ойнап жүр* – деген сөйлемдегі субъектің денотаттық статусын дәлдікпен айту қиын, ал олардың саны екіден аз еместігі анық деуге болады. Сол сияқты, сөйлем ішіндегі *кейбір, бірнешеу, бір-бірімен* сияқты белгісіздік есімдіктер де қажетті дәлдікті азайта түседі. Қайсыбір тілдесу кезінде мұндай дәлдік жеткілікті де болуы мүмкін. Ал басқа бір жағдайларда дәлдікті күшейту үшін қосымша айтылымдар қолданылады.

Енді тілдегі **жинақты қорытындылау** мәселесіне келсек, оған мысал болатын атау процесі. Мәселен, "мал" деген атаудың мағыналық құрылымында көптеген жан-жануарлардың жинақталған сыр-сипаты жатқаны мәлім.

Тілдік айтылымдардың жалпылау, жорамалдау және белгісіздікке, дәлдікке қатысты ерекшеліктері, жалпы алғанда, коммуникация нәтижесін жоққа шығармайды деуге болады. Ал тіпті, ондай ерекшеліктер айтылым мағынасын күңгірттендірсе, қалаған дәлдікке жету үшін айтылымды қосымша фразалармен толықтыруға әр уақытта мүмкіндік болады.

Модельдің модельденуші нысаннан айырмашылығы. Мұндай айырмашылық жоғарыда айтылған қасиеттерден қисынды түрде туындап тұр деуге болады. Тіпті адам қолымен жасалмаған модель, мысалы, су ағымының "үлкен иіріміне" сай келетін өлшемдерін қайсыбір "кіші иірім" өлшемдерімен салыстыратын болсақ, біріншіден, ол өз өлшемімен, қоршаған ортаға әсер ету күшімен және т.б. айырым табады деуге болады.

Сол сияқты, тілдік шығармалардың да өзінде көрініс табатын ақиқат шындықтан айырмашылығы болатыны сөзсіз.

Модельдің сапасы. Модельдер, модельденуші нысанды түпнұсқаға барабар етіп және оны толық қамту жағынан әртүрлі дәрежедегі дәлдікпен қайталайды. Модельдің сапасын анықтайтын мұндай дәлдіктер объективті факторлардан басқа субъективті факторларға да байланысты болып келеді. Субъективті факторларға модель құрастырушылардың білім деңгейі, қабілеті, тәсілқойлығы және қоршаған әлемді меңгеруге қажетті білім дәрежесін және т.б. жатқызуға болады. Бірақ "жаман" модельдердің өмірде кездесуінен, жалпы модельдеу идеясына қатысты күмән туғызу және бұл зерттеу бағытын толығымен қолдамауға болмайтыны анық.

Мәтіндер шындықты әртүрлі барабарлықпен және әртүрлі толықтық дәрежеде сипаттайды. Кейбір жағдайда, мысалы, әдейі таратылған жалған ақпарат негізінде құрылған модельде (шығармада, мәтінде) ақиқат шындыққа қарама-қарсы мәлімет болуы мүмкін. Бірақ бұған қарап әлемді мәтін арқылы модельдеу тезисіне күмән туғызуға болмайтындығы шындық.

Модель құрастырушылардың белсенділігі (активтігі). Модель, ол әр уақытта жеке адамның немесе ұжымның белсенді түрдегі интеллектуалды-физикалық қызметінің нәтижесі. Мәтін авторының ақиқат шындықты модельдеу процесінде тілдегі аспектілердің ең маңыздылары тілдесу мақсатына релевантты (аса қажетті) деп саналып, белсенді түрде таңдалып алынады. Осы сияқты хабарламаның мақсатын дұрыс таңдай білу де өзінше бір белсенділік пен мақсаттылықтың нәтижесі деуге болады.

Сонымен, ақиқат шындықтың мәтін арқылы белсенді және саналы түрде құрылған моделін адам әрі өзі туындатады, әрі өзі қабылдайды. Соңғы жағдайда, яғни адамның модельді қабылдауы тілдік тетік арқылы автоматты түрде іске асады.

Модельдеу жағдаятындағы мақсаттылық. Ақиқат болмысты тану ретінде құрастырылған модель белгілі бір әлеуметтік маңызды мақсатпен жасалады. Мысалы, ондай мақсатқа жататындар – адам қызметінің әр аясы бойынша жасалған модельдердің көмегімен анықталатын модельденуші нысан жайлы немесе табиғаттың неше түрлі құбылыстары жайлы мәліметтерді тіршілік қажеттігіне пайдалануды айтуға болады.

Тілдік шығармалар арқылы болмысты модельдеудегі әлеуметтік маңызды мақсаттарға жататындар: шындықтың әртүрлі аясындағы адамдар қызметін іске асыруға қажетті коммуникация мәселесін шешу. Қарым-қатынас кезінде байқалатын мақсаттылық принципін бұзу ережеге сай келмейтін ерекшелік болып саналады. Осы сияқты мақсаттылық шартын бұзушылық фактілері сөйлеу актының нәтижелі болуына кедергі жасайды деп саналады.

Жоғарыда келтірілген сәйкестіктердің негізгі мақсаты – тілдік шығармалар негізінде, яғни мәтін арқылы ақиқат әлемін модельдеуге болатындығына көз жеткізу. Бұл жағдайлар сөйлеу қызметі түрлерін модельдеу проблемасына байланысты туындайтын қиыншылықтарды түсіну үшін аса маңызды. Яғни бұл жердегі модельдеуге қатысты қиыншылықтар тіл қызметінің кибернетикалық моделін құруға қатысты мәселелер.

Сонымен, шағын пәндік аяға қатысты "сөйлеу тәртібі" (речевое поведение) жүйесін модельдеу тәжірибелері тілдік шығармалар негізінде әлемді модельдеуге болатындығы туралы пікір-талас теориялық жағынан қолдау тапты деуге негіз бар.

2.2 Тілдік тетікті модельдеу

Егер әр түрлі "сөйлеу тәртібінің" (речевое поведение) және оның міндетті түрдегі компоненті болып саналатын мәтіннің негіздерінде қайсыбір ортақтық болмаса, адам арасындағы тілдік қарым-қатынас та іске аспаған болар еді. Мұндай ортақтық әр нақты тіл үшін өзіне ғана тән тілдік механизмдер негізінде жасалады. Басқаша айтқанда, аталған ортақтық нақты тілдің грамматикасы мен сол тілдің сөздігі арқылы көрініс табады. Тілдік тетікті, барлық аспектілері бойынша, белгілі тәртіппен құрылған және анық түрде қызмет ететін тілдік жүйе деп те түсінуге болады. Мұндай шарт орын алмаған жағдайда, тілдік тетік адамдардың тілдесу қызметінің құралы бола алмай, адамдар бір-бірімен түсінісе алмас еді. Тілдің құрылымы қарапайым түрде болғанымен, қайсыбір алдын ала қабылданған болжамсыз тілдік әрекеттердің моделін жасау мүмкін болмас еді.

Автоматты түрде сөзбе-сөз аудару жүйесінің алғашқы тәжірибелерінде мынадай мәліметтер қамтылған: екі тілдік сөздік, орфографиялық қателерді түзету редакторы, мәтіндік негіз сөзді бөліп алу ережесі, аффикстер мен флексиялардың немесе лексемалар тобының сөзтұлғалық парадигмалардың (септелу, жіктелу үлгісінің) грамматикалық мағыналарының жиыны.

Сөйлеу тәртібінің ең "күшті" моделі өз құрамында тілдік жүйенің барлық аспектілерін қайта туындата алатын тілдік тетіктің "күшті" моделінің болуын қажет етеді. Жәй жағдайдағы тіл құрылымын лингвистикалық сипаттау мүмкіндігін интеллектуалды-құрылымдық модель тобына жатқызуға болады. Бұл модельдер көптеген әлеуметті-мағыналық мақсаттар үшін де өте тиімді деп саналады. Мысалы, олар ана тілі мен шет тілін оқытуда, аударма мақсатында, мәтінді редакторлық өңдеулерде және т.б. мақсаттар үшін тиімді деуге болады. Әрине, аталған модель интуицияны, тапқырлықты, болжам жасауды, тәжірибе нәтижелерін және т.б. кең қолдана білетін интеллекті (ойлы) адамға арналған. Сонымен бірге тілдік модель адам баласының "сөйлеу тәртібінің" қандай да болса түрлерін қайта туындата алатын кибернетикалық жүйенің болуын талап етеді. Сондықтан ондай модель адамның ойлау қасиетінің ерекшеліктерін "игерген" электрондық "интеллектке" бағышталған деуге болады.

Енді тілді модельдеудегі алда тұрған міндеттер мен кездесетін қиындықтар қандай болмақ, соған тоқталайық.

Адам баласы тілдік шығармалар арқылы белсенді және саналы түрде ақиқат болмысты модельдеуге әрекет етеді. Ол үшін тілдік тетікті (тілдік жүйені) автоматты түрде туындату мен қабылдаудың жолдарын зерттейді. Мұндай жағдайда "тілді" әдеттегі кибернетикалық "қара жәшік" тәріздес деп ұғыну қажет болады. Тілдің

ақиқаттығын және оның адамдар үшін қызметінің маңыздылығын біз жақсы білеміз. Бірақ тілді зерттеуші ғалымдар мен тіл арқылы қарым-қатынас жасаушылар үшін оның ішкі сыр-сипаты жұмбақ секілді – жасырын сипатта. Сондықтан тілдік коммуникацияға қатысушыларға тек қана "қара жәшікке" енетін және одан шығатын ақпарат қана белгілі деуге болады. Мәселен, мәтін авторына оның оқырманға (қабылдаушыға) жолдайтын ақпараттың мән-мағынасы (смысл) ғана белгілі делік. Ал ақпаратты қабылдау жағдайында керісінше, адресат қолында бар мәтінді меңгеріп, оның мағынасын түсіне білуі қажет. Сонда зерттеуші рөлін – үшінші реттегі бақылаушы ғана атқарады да, оның тілдік құрылым мен қызметін тікелей бақылауға мүмкіндігі болмайды. Зерттеуші немесе бақылаушы өз болжамдарын тілдік нәтижелерді талдап, зерттеу арқылы біле алады.

Тіл құрылымының бақылауға жатпайтын қасиеттері тілдік тетікті модельдеу ісіндегі негізгі қиындықты туғызатыны белгілі.

Теориялық тіл біліміндегі идеялар мен қолданбалы лингвистика саласындағы "сөйлеу тәртібінің" кибернетикалық моделін құру тәжірибелер нәтижелері жайлы былай деуге болар еді. Тіл тетігін туындату мен оны қабылдау, олардың берілістерімен, мақсаттарымен және амалдарымен айрықшаланады. Бұл айырмашылық мәтіндерді талдау (қабылдау, тану, түсіну) мен жинақтау (туындату) модельдерінен жақсы байқалады. Лингвистикалық құбылыстар және соның негізінде құрылатын теориялар аталған айырмашылықтарды бөлшектеп қарастырудың негізінде ғана түсініктеме алады. Мысал ретінде, лингвистикада көп айтылып жүрген "пресуппозиция" ұғымын алуға болады. Мәтінді қабылдау жағдаятында ол басқаша түсінік алады. *"Мұрат тұмаумен ауырып қалды"; "Әлия қайтып келеді"* деген фразаларды құрудың алдында, автор *Мұрат пен Әлияның* өмір сүріп жатқанын (экзистенциалды пресуппозиция), *Мұраттың* бұрын ондай ауруға шалдықпағанын (ықтималды контрастирлену пресуппозициясы) және *Әлияның* бір жаққа кеткенін (абсолютті контрастирлену пресуппозиция) білетіндігі негіз болып отыр. Сонымен бұл түсінік мәтін авторының ұстанған көзқарасынан туындап отыр. Яғни кез келген мәтінге түсініктеме беретін кибернетикалық жүйе мәтіннің мағыналы мазмұнын аша білуі керек.

Ал мәтіннің мазмұны адресат-бақылаушыға арнайы арналмағандықтан, оған мағыналы қорытынды жасаудың қиындығы артады.

Модельдеудің түпкі мақсатын тағайындай білу де аса маңызды іс. Бұл мәселеде осалдық танытуға болмайды. Себебі ондай жағдайда коммуникацияның түпкі мақсаты толық орындалмай, модель құру әрекеті тек қана зерттеушінің тілдік білім дәрежесін модельдеумен ауысуы мүмкін. Ал тілші қызметін модельдеу компьютерлік лингвистика саласының дербес мәселесі екенін ескеру қажет.

Тілдік тетікті модельдеудің келесі проблемасының түпкі мақсаты – мәтінді туындату мәселесі. Ал бұл, жоғарыда аталғандай, болмыстың қайсыбір үзіндісін ғана қайта туындататын "бірінші" реттегі моделі деуге болады. Осыған байланысты мағынаны белгілеу мен анықтау, мазмұн мен тұрпаттың сәйкестігі, мазмұн мен ақиқат шындықтың сәйкестігін анықтау маңызды мәселеге айналып, референция ұғымы алдыңғы қатардағы проблемаға шығып отыр. Осымен бірге семантика мен референция арасындағы байланысты нақты түрде сипаттау қажет болуда. Әртүрлі жағдаятқа қатысты мұндай байланысты сипаттауда сөздіктің атқаратын рөлі зор екендігі белгілі. Шындығында да, табиғи жағдайдағы сөйлеу қызметіне қатысты қайсыбір тілді сөздіктен тыс қабылдау (елестету) мүмкін емес те. Ана тілін үйреніп жүрген жас бала немесе шет тілін меңгеру жолындағы ересек адамдардың бәрі де

тілді абстракты түрде емес, олар қажетті тілдік ақпаратты мағынасы анық болып келетін нақты сөздер арқылы қабылдайды. Тілдік тетік туралы бірден-бір дұрыс түсінік, тек қана сөздердің қолдану мағынасына қатысты білім қорын жинақтау арқылы пайда болады. Бұл тұжырым сөз тіркестеріне байланысты да орын алады.

Әрбір сөздік негіз сөздер (лексема) және сөз тіркестері топтамаларының көрінісі ретінде түзілетіні белгілі. Егер құрастыратын сөздік түсіндірме сөздік сипатында болса, онда сөздіктің мақала мәтіні сөз не сөз тіркестерінің мағыналық бірлігі негізінде қарастырылады. Сондықтан бұл жұмыстар, яғни тілдік механизмге лексикографиялық талдау жүргізудің өзіндік қиындықтары да баршылық.

Сонымен, сөздер негізінде мағыналық бірліктердің көрініс табуы тілдік таңбалар жиыны жайлы тек болжам ғана деуге болады. Солай бола тұра бұл болжамның ықтималдығы өте жоғары дәрежеде.

Аталған және басқа лексикографиялық талдау жүргізудің қиындықтарға қатысты түрлерін алдын ала білу автоматты сөздік құрастыруда аса маңызды мәлімет болып саналады.

Келесі айтпақшы ойымыз – ол семантика құрылымын жүйелі түрде суреттеу тұстарында сөздер мен айтылымдардың омонимдік және синонимдік қатарларын ажырату ережелерін анықтау да маңызды. Әсіресе мұндай маңыздылықтың дәрежесі жасанды интеллектің ойлау әрекетіне қатысты эксперттік жүйелер құруда барынша арта түседі. Ал мұндай жұмыстың табиғи тіл негізінде іске асатыны белгілі.

Қандай да болсын модель қызметінің тиімділігін тексеруге сөйлеу қызметінің кибернетикалық модельдерін құрудың болымды нәтижелерінің айтарлықтай көмегі барын ескеру қажет. Сол сияқты, тілді модельдеуге қатысты автоматтандыру мәселелерін шешуде мынандай жайт айқындалуда. Дәстүрлі зерттеулердегі тілдік сипаттамалардың жоғары сапалығына қарамай, оларды сол күйінде автоматтандыру жұмыстарына пайдалану мүмкін еместігі мәлім болды.

Тілдің математикалық моделі дегеніміз, ғалым Г.С.Цейтиннің анықтамасы бойынша, ақиқат тілдің қасиеттеріне жақын келетін қайсыбір математикалық "құрылғы" [Цейтин, 1959]. Егер тілдің математикалық моделін Цейтинше түсінсек, онда автоматтандыруға қатысты тілді сипаттауға ең алғаш әсерін тигізетін нақты ғылымдар әдістері мен формалдау әдістері болмақ.

Математикалық құрылғылардың түрлеріне қарай, тілдік тетіктің де әртүрлі математикалық модельдерін құруға болады. Мысалы, сөйлемдер тәуелділігінің дәл құрылымын сипаттауға бағытталған математикалық модельді құру үшін графтардың математикалық теориясы аппаратын және алгебраның, математикалық логиканың әдіс-тәсілдері қолданыс тапты. Зерттеушілердің тілді математикалық модельдеуге қатысты жасаған іс-әрекеттері, оларды тілдік тетік табиғатына жаңаша көзқараспен қарауға ынталандырды деуге болады. Тілдік математикалық модельдер "сөйлеу тәртібінің" кибернетикалық модельдерінде қолданыла бастады. Мысалы, алғашында олар машиналық аударма жасау жүйесінде пайдаланылғаны мәлім. Бұл модельдерді іс жүзінде пайдалану кезінде олардың әлсіз тұстары да байқала бастады. Мәселен, кейбір сапалы құрылған математикалық конструкциялардың өздері табиғи тіл құбылыстарының сипаттамаларын жан-жақты қамти алмады. Әрбір модельге тән қасиет – модельденуші нысанды жинақтап қорыту және оны жеңіл түрде сипаттау. Бірақ бұл нәтижелер барлық тілдік құбылыстарға ортақ емес. Басқаша айтқанда, іске қосылған модельдерде сәйкес қасиет толық сипатта болмады және олардағы модельденуші нысандардың қасиеттері дұрыс түрде қайта туындатылмайды.

Автоматты модельдеу әдісінің даму сипаты лингвистикалық және математикалық модельдерді өзара сәйкестендіру арқылы жүргізуді ең бір дұрыс жол деп тапты. Бұл шешімнің аса маңыздылығын қолданбалы және құрылымдық лингвистика саласының даму қарқынынан да аңғаруға болады.

Тіл ғылымы саласындағы модельдеудің объективті қиындықтарының бірі, ол тілдік тетіктің статикалық қалпынан гөрі, тілдік шығарманы туындату мен қабылдаудағы динамикалық (даму) процесін қайта туындататын толық және балама сипаттағы лингвистикалық модельдің болмауы.

Тілдік тетіктің жұмсалымдық қызметінде динамикалық процестің бақылаудан тыс жатқан құбылысқа қатысты болуы да оның моделін жасауға қиындық тудырды. Мәселен, "тілдік тетіктің жұмсалымдық қызметі" ұғымында тілдік тұлғалар мен құрылғылардың қолдану жүйешесін (подсистемы) ескеру қажет екені дәлелденді. Ал автоматтанған тілдік мәселелерге байланысты мұндай жүйешелердің қолдану сипатының зерттелуі әлі де қанағаттандырғысыз болып отыр.

Қолданбалы лингвистиканың арнайы мәселесінің бірі – сөйлеу тілін қабылдау мен сөйлеу әрекетін туындататын кибернетикалық жүйелерді модельдеу проблемасы. Мұндай модельдерді құрудың алдында фонологияның құрылымдық және сөйлеу тілін туындату мен қабылдаудың динамикалық моделін құру мәселелері жете зерттелуі қажет.

Әдетте кибернетикалық жүйелерде тілдік тетік синхронды аспектіде модельденеді. Бірақ тілдің синхронды бөлігі жайлы ұғым – барынша шартты. Оның себебі, тіл, әр алуан әрекет өрісіне қызмет етуімен бірге, басқа тілдермен де ұштасып, және олардан территориялық алшақтықта тұрып та үздіксіз өзгеріске ұшырай береді. Шындығында, тіл грамматикасының өзгеруімен бірге тілдік қор да жаңа сөздер мен сөз тіркестері арқылы полисемия мен омонимияның даму негізінде және т.б. арқылы үнемі ұлғайып отырады.

Жоғары дәрежелі сәйкестік сипаттағы модельдердің аталған инновацияларын, яғни тілдік қордың жаңаруларын ескеретін мүмкіндігі болуы қажет. Мұндай шартты қанағаттандырмайтын тілдік тетік мүмкіндік моделі шектеулі модельдер қатарына жатады. Тілдің математикалық моделі кибернетикалық жүйелерді жабдықтау компонентінің бірі ғана болып есептеледі. Ал "тілдік тәртіпке" қатысты модельдерге байланысты құрылымдық лингвистикалық модельді құрастыру жолдарын қарастыру керек. Мұндай модель тілдік жүйе қызметін қайта туындату процесінен басқа, тілдік жүйенің құрылымдық жағын да қайта туындата алады деп саналады.

Әдеттегі лингвистикалық сипаттамаларға қарағанда, аталған модель, өзінің дәлдік (нақтылық), сәйкестік, қарапайымдық, жинақтылық, қайшылықсыздық және толықтық талаптарын қанағаттандыра алу мүмкіндігімен айырым табады. Сонымен бірге модельдің сәйкестік пен толықтық қасиеттерінің шектеулігі жайлы арнайы түрде зерттеулер жүргізілуі керек.

2.3 Тіл модельдерінің түрлері

2.3.1 Лингвистикадағы дедуктивті әдіс

Қай ғылым саласы болмасын, оның дерексізденуге жүгінетіні белгілі. Яғни зерттеу нысанына тән зат не құбылыстың қасиеттері мен олардың арасында туындайтын қатынастардың кейбіреулерін ой қызметінен тыс зерттеу. Бірақ ондай

дерексізденуді таңдау жолы әр түрлі болуы мүмкін. Қайсыбір ғылымдарда, мысалы, геология, ботаника, химия, физика салаларында табиғатта кездесетін нақты құбылыстар деректерін жинақтау әдісі қолданылады. Басқаша айтқанда, бұл әдіс жалқыдан жалпыға қарай бетбұрыс арқылы қорытынды жасауға сүйенеді. Ал басқаларында, мысалы, математика мен логикада алдын-ала белгілі деп саналатын (априорно) жалпы абстрактылық шындақтан жекелік фактіге қатысты түйін жасалады. Қорыта келгенде, бұл аталған зерттеулер ой қызметі негізінде тұжырым жасалатын – дедукция, яғни саралау, талдау арқылы жалпыдан жалқыға немесе бөлшекке көшіп нәтиже шығарудың және индукция – жинақтау арқылы жалқы, жеке жағдайлардан жалпы қорытындылар шығарудың ойлау тәсілдері деп аталып жүр. Әрине, таза пәк дедукция әдісіне не индукция әдісіне сүйенетін ғылым саласы жоқ деуге болады. Мәселен, индукция әдісін қолданатын химия мен физика қазіргі көрінісінде елеулі дәрежеде дедуктивті бөліктен тұрады, керісінше математика ғылымында индукция элементі елеулі рөл атқарады [Ф. де Соссюр, 1933, 111-б.]. Аталған жайттарға қарамай, зерттеу барысындағы әдістердің басымшылығының негізінде ғылым салаларын – *дедуктивті* және *индуктивті* ғылымдар деп ажыратып жүр.

Олай болса, тіл білімі қай ғылым саласына жатады? Бір жағынан алғанда, тіл білімі өзінің табиғатына сай, нақты тілдердің нақты деректерін сипаттайтын құндылығына қарай, оны индуктивті ғылым деуге болады. Ал екінші жағынан, тілшілердің зерттеу нысанына айналып отырған сөйлеу актілерінің шексіз көп жиыны екендігін ескерсек, онда индукция тәсілінде орын алатын жалқыдан жалпыға қарай қарастыру тіл ғылымының негізгі ұғымдарын жалпылама тұрғыда тұжырымдау аталған әдіс арқылы мүмкін де емес.

Бұдан туындайтын қорытынды – тілшілер өз зерттеулерінде индуктивті әдіспен бірге дедуктивті әдісті де қатар қолдануы қажет. Осындай әрекеттің негізінде нақты тілді талдау арқылы қол жеткен деректерді ұғынуға қажетті жалпы ұғымдар жүйесін құруға болады.

Тіл білімі өзінің дедуктивті бөлігінде, сірә, логика немесе математика ғылымдарына ұқсас құрылған тәрізді. Яғни қайсыбір алғашқылар ретіндегі және айқындалуы қиын, сан жағынан шағын болып келетін терминдер бөлектеліп алынады да, ал басқа терминдер осы алғашқылар арқылы анықталады. Ол үшін терминдердің өзара байланыстылық сипаты жайлы пайымдаулар нақты түрде тұжырымдалып, аксиома түрінде дәлелсіз қабылдануы керек. Ал басқа пайымдаулардың барлығы бірдей дәлелденуі қажет, яғни олар бұрын тұжырымдалған пайымдауларға саяды.

Математикалық лингвистика айналасында болып жатқан көптеген келіспеушіліктердің бәрі тіл ғылымын математика құрылымына сәйкестендіру мүмкіндігі жайлы екені мәлім. Бұл жөнінен әзірше бір пікірге келу қиын, себебі нақты тілдердің грамматикаларында кездесетін барынша қарама-қайшылыққа толы тіл ақиқаттығы фактілерінің әр алуандығын дедуктивтік теорияның шеңберіне сыйғызу оңай емес. Бірақ мұндай қорытындылау зерттеуші-ғалымдардың пессимистік пікіріне жол бермеуі керек. Себебі қазіргі замандағы ғылым ақиқат шындықтың күрделі құбылыстарын танудың зор құралы болып саналатын модельдеу әдісіне ие болып отыр.

Сонымен, Соссюрше айтқанда, модельдеу әдісін, ең алдымен, лингвистиканың "ішкі бөлігі" деп аталатын жағына қолдану мақсатқа сай келеді деуге болады. Әрине, мұндай жағдайда модельдердің нақты тілдерге қатысы жоқ ерекшеліктерін қарастырған жөн болады. Әрі қарайғы зерттеулерде жеке тілдер топтарын, мүмкін, тіпті, нақты тілді де модельдеуге мүмкіндік туар.

Қорыта айтқанда, тілді модельдеу арқылы нақтылы ақиқат шындықтың деректеріне өте-мөте жақын жатқан тізбектердің дерексіз сызбасы құрастырылады [Андреев, Зиндер, 1959, 18-б]. Сонымен, тілді модельдеу дегеніміз ол – әдіс. Бұл әдіс бойынша зерттеуші нақты бір тілге қатысты жалпы қасиеттер арқылы тіл құрылымын "абстрактылы семиотикалық жүйе" деген ғылыми болжамға сүйене отырып, осыдан туындайтын нақты тілдік деректерге қатынасының салдары анықталады.

2.3.2 Модель және оның интерпретациясы

Тіл моделінің жалпы көрінісі. Модельдеу қажеттігі зерттеу нысанды тікелей байқауға (яғни қолмен ұстап, көзбен көруге) мүмкін болмайтын жағдайлардан туындайды. Мысалы, физиология атты ғылым саласында жоғары нерв талшықтарының қызметін зерттеуде немесе атомдық физика саласындағы атом ішіндегі процестерді қарастыруда модельдеу әдісінсіз ғылыми нәтижелерге жету мүмкін емес.

Ал аталған физиолог және физик мамандарымен салыстырғанда тілші-ғалымдардың да еш айырмашылығы жоқ. Себебі оның зерттеу нысаны да ішкі құрылымы жағынан күрделі келетін мәтін не сөйлеу қызметіндегі дыбыстық элементтер болып табылады. Бұл аталғандар да тікелей бақылауға ыңғайлы бола бермейді. Сондықтан бұл өзінше бір бөлек әдіс ретінде көптеген ғылымдар саласында қолданылады.

Сонымен, модель дегеніміз зерттеу нысанының қасиеттерін бойына сақтай алатын және түпнұсқаның орнына жүретін, таным процесінің қажеттігінен туындайтын материалдық не абстрактылық нысан.

Модельді құру үшін, ең алдымен, нақты ғылым саласы бойынша жинақталған әр алуан ұғымдардың ішінен "алғашқылары" (бастапқылары) деп санауға ыңғайлы келетін кейбіреулері таңдалып алынады. Алғашқы ұғымдар арасындағы қайсыбір қатынастар постулат (дәлелдеуді керек етпейтін ұғымдар) ретінде белгіленуі қажет. Ал басқа тұжырымдамалар алғашқы ұғымдардың терминдері негізінде қатал дедуктивті тұрғыда туындауы керек.

Бұл түсініктегі "модель" жүйе түріндегі тіл бөлігі емес, ол қайсыбір гипотезалық (болжаммен алынған) ғылыми құрылым, белгілібір құрылғы [Шаумян, 1960, 150-б.].

Модельді, яғни элементтердің табиғатына тәуелсіз болатын қайсыбір абстракты нысанды және оның лингвистикалық интерпретациясын (түсінігін) қатал түрде ажырата білген жөн. Сондықтан әрі қарайға баяндауымызда модель ұғымы үшін математика саласында ұстаным ететін терминдерді, ал модельдің түсінігі ретінде, мүмкіндігінше, тіл білімінде қабылданған терминдерді пайдаланамыз. Бір ғана модельдің лингвистикалық түсінігі бірнешеу болуы мүмкін. Мысалы, *фонологиялық* және *грамматикалық*.

Модель мен оның түсінігі арасындағы өзара қатынастар аса күрделі және әр алуан деуге болады. Жоғарыда айтқандай, модель – дерексіз құрылғы, ал оның түсінігі жайлы проблема болса, Б.Рассельше айтқанда: "... проблема интерпретации есть проблема подстановки чего-либо определенного" [Рассел , 1957, 276-б.] немесе модельдің орнына қойылатын нақты "белгісіз бір" мәселе. Ал тіл ғылымы саласындағы зерттеулерде кездесетін "модель" терминін ақиқат лингвистикалық нысан деп ұғынған жөн.

Лингвистикалық модельдің қайсысының болса да түсінігі алдын ала анықталмайынша, ол модель шындыққа жанаспайтын жалған жағдаймен бірдей. Шындығында да, логикалық тұрғыда дедуктивті жолмен алынған дәлел қаншалықты мінсіз болғанымен модель арқылы сипатталған тілдің ақиқат қасиеттері еш мәлімет бере алмайды. Ал егер, түпнұсқа мен модельдің арасындағы сәйкестіктер сай келіп жатса, модельді дәлелді деп санауымызға болады [Зиновьев, Ревзин, 1960, 83-85-бб.]. Мұндай сәйкестіктің белгісі тек тәжірибе жүзінде анықталады. Сондықтан модельдер теориясы үшін тіл білімінің "машиналық аудармаға" қатысты және т.б. да тәжірибелік тұстарының маңызы зор.

Енді лингвистикалық модельдердің алғашқы ұғымдарының сипаттамаларын сөз етелік.

Белгілі бір континуум (үздіксіз) сипатындағы акустикалық фонетикаға қарағанда, модельдер теориясы әр уақытта **дискретті (бөлікті) бірліктерден** тұрады. Сол себепті де көптеген ғалымдар лингвистиканы **"дискретті математика"** саласына жатқызып жүр. Модельдер теориясын таза лингвистикалық теория деп қарастыруға болады. Осыған қарамай, кез келген лингвистикалық **модельдердің дискреттілігін (бөліктілігін)** мойындау аса маңызды мәселе.

Барлық модельдерде алғашқы ұстаным ретінде қайсыбір элемент ұғымы алынады. Модельдеуге қатысты тіл деңгейіне қарай, бұл элемент әртүрлі болуы мүмкін. Мысалы, фонетикалық деңгейде ондай элемент бөлек бір дыбыс немесе синтаксистік деңгейде – сөзформа (сөзтұлға) болуы мүмкін.

Модельдерге қатысты келесі бір ұғым – **"кортеж"** ұғымы. Бұл ұғымға кез келген элементтердің реттелген тізбегін жатқызады. Мысалы, $A=x_1x_2 \ldots x_i \ldots x_n$.

Қарастырып отырған тіл аясына қатысты немесе қатыссыз болуына қарай, **"кортеждер"** бірінші кластық – **белгі қойылған** (отмеченные) және екінші кластық – **белгі қойылмаған** (не отмеченные), деп екіге ажыратылады. **"Элемент"** ұғымы тәріздес **"кортеж"** ұғымы да тілдің зерттелу деңгейіне байланысты әртүрлі түсінік алады. Мысалы, фонологиялық деңгейде **"белгі қойылған кортеж"** деп **"фонетикалық сөз"** ұғымы алынатын болса, ал синтаксистік деңгейде **"кортеж"** ұғымы фраза ретінде түсіндіріледі де, **"белгі қойылған фразалар"** деген терминмен аталады [Ревзин, 1962, 11-б.].

Модель элементтерінің саны мен **"белгі қойылған кортеж"** ұзындығының шекті не шексіз болуына қарай, тіл модельдерін негізгі төрт түрге бөліп қарастырады. Ол модельдердің әрбіреуі зерттеуші ғалымға айырықша сипаттағы логикалық мәселелерін қояды. Біздің алғашында қарастыратын моделіміз – элемент саны жағынан шекті болып келеді де, ал кортеж ұзындығына ондай шек қойылмайды.

2.3.3 Аналитикалық және синтетикалық модельдер

Тіл модельдері **аналитикалық** және **синтетикалық** деп екіге ажыратылатыны белгілі. Бұл модельдің логикалық құрылымына қарай, біз ондағы нысандар жиынына бірнеше формалды амалдар түрін қолдана аламыз. Қолданылатын амалдардың түрлеріне байланысты барлық зерттеу нысандарын үшке бөліп қарастыруға болады:

1) амал қолдану кезінде сырт жағдайдан туындайтын және белгілі бір амал ғана қолдануға болатын **бастапқы нысандар**;

2) амал қолдану кезінде туындайтын және оларға да амал қолдануға болатын **аралық нысандар**;

3) амал қолдану кезінде туындайтын, бірақ оларға амал қолдануға болмайтын **қорытынды нысандар**.

1-ші анықтама. *Егер "А" моделі үшін бастапқы болып саналатын нысандар "В" моделі үшін қорытынды нысан болса, онда "А" моделі "В" моделіне қарағанда кері модель деп аталады; "В" моделінің бастапқы нысаны, "А" үшін қорытынды нысан болады; "А" моделіне қарағанда, оның кері моделін құруды, шартты түрде, "А" моделінің "айналымы" (обращение) деп атайды.*

2-ші анықтама. *Аналитикалық (немесе талдау) модельге қарағандағы кері модельді – синтетикалық (жинақтау) моделі деп атайды; егер модель құруда* ***белгі қойылған кортеж*** *жиынына сүйенетін болсақ – **аналитикалық** (талдау) **моделі**, ал егер қайсыбір амалдар негізінде **белгі қойылған кортеж** пайда болса, – **синтетикалық** (жинақтау немесе туындату) **моделі** деп атайды.*

Модельдердің бұл екі түрі (аналитикалық және синтетикалық) лингвистикалық сипаттаудың екі түріне сәйкес келеді. Олардың біріншісі – сөйлеу әрекетінен тілдік жүйеге қарай бағытталған сипат, екіншісі, керісінше, тілдік жүйеден тілдік дерекке қарай бағытталған сипат. Бұлар жағдаятқа байланысты коммуникацияның қабылдау актісіне сәйкес талдау мен сөйлеу актісіне сәйкес жинақтау сипатында болады.

2.3.4. Лингвистикалық модельдердің негізгі қасиеттері

Лингвистикалық модельдердің негізгі деп саналатын бес түрлі қасиеті анықталады:

1. Модельдеуге жататын құбылысқа тән қасиет – оның физикалық табиғатына қатысынсыз, тек құрылымдық (жұмсалымдық) сипатына ғана қатысты болуы.

Тіл табиғатының Соссюрше түсінігі бойынша, мұндай негізгі белгілер (жұмсалымдық немесе құрылымдық) тіл қасиетінде де бар.

2. Модель әр уақытта өзінің түпнұсқа түріндегі шындығынан да артық күйде, яғни дәріптелген (идеалды) сипатта түсінік алады. Мысалы, қайсыбір айтылым кезіндегі сөйлем мүшелерінің толық болмауына қарамай, тыңдаушы жоқ сөздің орнын "өз ойымен толықтырып" сөйлем мағынасын түсіне алады. Сол сияқты, әдетте, мәтін "ұзындығы" не оның ішіндегі сөйлемдер саны шекті болса да, теориялық жағынан алғанда оларға ешбір шек қойылмайды. Яғни грамматикалық ережелер бойынша да сөйлем ұзындығына шек қою жоқ деген сөз. Сондықтан да лингвистикалық модельдерге қатысты сөз болғанда, грамматиканың мүмкіндігі де дәріптеледі, яғни "мүмкіншілікті асыра сілтеу" жағдайы да кездесуі мүмкін.

Осы себептен, грамматиканың ақиқат мүмкіндіктеріне де шек қойылмайды. Мысалы, сөйлем ұзындығын шектемей қарастыратын жағдайдағы грамматика әрі "өткір", әрі "терең" сипатқа ие болады. Осылайша мүмкіндікті кең түрде пайдалану сипаттауды қарапайым және жеңіл түрде іске асыруға жағдай жасайды. Мысалы, сөйлем ұзындығын шектеуге қатысты тұжырымды жеңіл немесе аса күрделі ереже түрлерімен беру мүмкін бола бермейтінін айқын.

3. Әдетте модель ақиқат нысандар ұғымдарымен әрекет жасамайды, ол тек белгілі құрылғы арқылы, яғни дәріптелген нысандар ұғымымен әрекет етеді. Соңғы аталған нысандар тәжірибелік деректерден тікелей және бірізді ұғымнан туындамайды. Олар көптеген байқау жиынтығы мен зерттеушінің көрегендік сезімталдығы негізіндегі жалпы болжам жасаудың "еркіндік" сипатында құрылады.

Кез келген модель – белгілі математикалық аппарат арқылы жасалған болжамның логикалық қорытындысының құрылғысы.

Лингвистикалық құрылымдардың фонетика мен семантика құбылыстарын танып-білуге арналса да, оның мұндай субстанцияларға тікелей қатысы жоқ деп түсіну ең дұрыс шешім болмақ. Бірақ бұл тұжырымнан мағына лингвистикалық теорияның нысаны бола алмайды, деген пікір туындамауы керек. Себебі, тілдің негізгі қызметі – мән-мағынаны тарату. Ал семантикалық модель құру – тілді толық сипаттаудың ең қажет бөлігі.

4. Кез келген модель, оның ішінде лингвистикалық модель формалды пішінде болуы қажет. Модельді "формалды" деп атау үшін зерттеуге тиісті бірліктер ішінен біреуі ғана (ең бастапқысы) анық және бірмағыналы түрде берілуі тиіс. Сонымен бірге, ондай бастапқы бірліктерді бір-бірімен тығыз байланыста ұстайтын тұжырымдар мен әрекет жасау ережелері де алдын ала айқындалуы керек.

Ең жақсы деп дәріптелген формалды модель "математикалық жүйе" деген ұғымға сай келуі керек. Сондықтан формалдылықтың мағынасы – математикалық дәлдік немесе бірізділік. Мұндай қасиеттер тіл табиғатына да тән.

Формалдылық, дәлдік, бірмәнділік – тіл теориясын пайымдаудың бірден-бір қасиеті. Әрбіреуі бөлек тұрғанда мұндай тілдік қасиет объективті тәжірибелік деректердің формалды теориясының болжамымен сәйкес келе бермейді. Теорияның дәлдігі әр уақытта тәжірибе нәтижесін дұрыстайды немесе терістейді.

5. Кез келген модель, соның ішінде лингвистикалық модель эксплонаторлық қасиетке ие болады деп саналады, яғни ондай модельдің тәжірибе нәтижесіне түсініктеме беру мүмкіншілігі бар деген сөз. Ал тілдік модель мұндай қасиетке ие болуы үшін, ең алдымен, осы кезге дейінгі түсініксіз деректердің анығына жетуге мүмкіндігі болуы керек. Сонымен бірге зерттеу нысанының бұрын ескерілмей жүрген жаңа мүмкіндіктеріне ғылыми болжам жасай алатын қасиеті болуы қажет.

2.3.5. Зерттеу модельдері

Зерттеушінің іс-әрекетіне еліктеп, оны қайталай алатын лингвистикалық модельдерді де құруға болатыны белгілі. Ондай модельдерді **зерттеу модельдері** деп атайды. Осындай модельдеу нәтижесінде мәтіннен тілдік жүйеге ауысуға жағдай жасалады да ондай жүйенің құрылымы жайлы құнды-құнды теориялық қорытындылар жасауға мүмкіндік туады.

Тілдік жүйе жеткілікті түрде танылуы үшін, яғни оның барлық сыр-сипаты зерттеліп болды деп шешім қабылдау үшін табиғи тілдің қарапайым бірліктері мен олардың топтары және әр тілдік деңгейдегі (семантикалық деңгейді қоса) түрлерінің тіркесімдік заңдылықтары толық танылуы қажет.

Зерттеу модельдері оларда пайдаланылатын бастапқы ақпараттың түр-тұрпатына қарай үш топқа бөліп қарастыруға болады.

Бірінші топтағы модельдерде бастапқы ақпарат ретінде тек мәтіннің өзі ғана пайдаланылады. Ал осы мәтіннен туындайтын тілдік жүйе, яғни ол мәтінді туындататын тілдік деректер тек осы мәтін ішінен алынады.

Екінші топқа қатысты модельдерде бастапқы ақпарат ретінде мәтінмен бірге осы тілдегі көптеген "дұрыс фразалардың" жиынтығы беріледі. Іс жүзінде, екінші топтағы модельді зерттеп дайындау кезінде зерттеуші қосымша бір "хабарлаушының" (информант) көмегін қажет етеді. Ол көмекші адам сынауға

ұсынылған әрбір фраза (айтылым) жайлы өз ойын білдіріп, оның дұрыс не бұрыстығын хабарлап отыруы қажет. Мұндай "хабарлаушы" ретінде тіл маманы да бола алады. Әрине, ол үшін "тілші-хабарлаушы" модельдеу тілін де жетік білуі қажет.

Ал енді үшінші топқа жататын модельдерде бастапқы тілдік ақпарат ретінде мәтін және **дұрыс фразалар жиынымен** бірге **семантикалық айырым белгілер жиыны** да алдын ала берілуі керек.

Бұл топтағы модельдеу әдісінде, тәжірибеде кездесетін фразалардың дұрыстығын және олардың мағыналық айырым-белгі жақтарын ажырататын **"хабаршының"** көмегі қажет болады.

Енді нақтылы бір тілдік процеске жататын – адам баласының **сөйлеу әрекетінің моделіне**, не басқаша айтқанда **"тіл меңгеру моделіне"** қысқаша тоқталайық.

Ең алдымен, **"тіл меңгеру"** терминінде қандай мән-мағына жатқанын ажырата білген жөн. Басқаша айтқанда, тілді меңгеру жағдаятында тілдің мағыналық жағы ескеріле ме жоқ па немесе мағына тіл құрылымының ішкі құрылысына қатысты ма жоқ па, деген сауалдарды алдын ала айқындап алуымыз керек болады. Қолданбалы тіл білімі үшін бұл жағдайды анықтау өте маңызды мәселе деп саналады.

Белгілі ғалым Л.Блумфильдтің ұйғаруынша, мағына тілдік құрылымнан тыс жатады және тілдік нақты мағына жағын ашу әр ғылым саласындағы мамандардың ісі деген түсінік туындайды [Блумфильд, 1968].

Мысалы, нақты ғылым саласының терминдерін жетік білмей тұрып-ақ, кейбір жағдайда тілдесушілер белгілі-бір тілде қарым-қатынас жасай алады деген сауал да туындауы мүмкін. Міне, осындай ой-пікірлердің негізінде американдық дескриптивизм ағымының өкілдері "мағына" тіл құрылымына қатысты емес деген тұжырымға да келген. Олардың ұйғаруынша, тілді меңгеру үшін айтушы (сөйлеуші) грамматикалық дұрыс фразаларды құра білсе, ал тыңдаушы сол фразаларды (сөйлемшелерді) дұрыс түсіне білсе жеткілікті деп қорытындылайды.

Шынына келгенде, грамматика мен мағына бір-біріне тығыз байланыста. Себебі, мағынасыз мәтіндер ертелі-кеш өзінің грамматикасын жоғалтуы, ал грамматика жағынан дұрыс құрылмаған мәтіндер көп ұзамай-ақ мағыналарынан айырылуы мүмкін. Сондықтан егер тіл қызметі адамдар арасындағы қатынас құралы екендігін ескерсек, онда "мағына" тіл құрылымына сөзсіз енеді. Олай болатын болса, тіл құрылымына мағынаның қандай түрлерінің қатысты екендігіне немесе олардың қайсылары тіл білімінің зерттеу нысаны болатынын сөз етелік.

Ғалымдар (И.А.Мельчук және т.б.) осы мәселеге байланысты мағынаны грамматикалық және грамматикалық емес, деп екіге ажыратады. Мағынаның грамматикалық болуы үшін сол мағына осы тілде кездескен түрпатында міндетті түрде қолданылуы қажет. Яғни айтылым кезінде осындай тілдік элементтің кездесуі әр уақытта бірдей грамматикалық мағынамен сәйкес келуімен қатар, өздері бөлек бір ірі топ құрып, мәтін ішінде де жиі қолданыста болады. Егерде кейбір мағынаның айтылымдарда кездесуі аса шартты болмай, тілде сирек қолданылса, ондай тілдік элемент грамматикалық емес мағына деп саналады. Мысалы, тілде заттың көптік не жекеше мағынасын білдіретін сөздердің грамматикалық мағынасы бір тілде сирек кездесуі мүмкін, соған қарай, ондай нақты мағына сол тілге **тән емес** деуге болады. Сол сияқты, орыс тілінің табиғатына тән "род" категориясының (муж., жен., сред.) грамматикалық мағыналары қазақ тіліне тән емес. Сондықтан әр тілге қатысты грамматикалық мағыналарды топтау, сұрыптау тіл білімінің зерттеу нысаны болғаны жөн. Жоғарыдағы айтылған, мағынаның тіл құрылымынан міндетті түрде орын алуына қарсы пікірдің болуын да ескеру қажет.

"Тіл меңгеру" ұғымына қандай мән-мағына берілуіне байланысты **сөйлеу қызметі моделінің** түрі екіге бөлінеді:

1) семантикасыз немесе грамматиканы игеруге "еліктеуші" таза синтаксистік модель. Яғни бұл сол тіл иесінің түсінушілік қабілеті мен грамматикалық дұрыс фраза құруына еліктеуші және фразаның мән-мағыналы болуы шартты емес деген ұйғарымға қатысты модельдер түрі;

2) тіл иесінің мағыналы сөйлемдерді түсінуімен бірге ондай сөйлемдерді дұрыс құруға еліктейтін семантикалық модельдер.

Модельдеу нысаны ретінде тіл қызметінің сөйлеу не қабылдау түріне қарай, тілдік модельдерді синтездеу (жинақтау) моделі және анализдеу (талдау) моделі деп ажыратылатыны жайлы аналитикалық және синтетикалық модельдерді сөз еткенде де айтылған болатын. Сол тақырыптағы көтерілген мәселелерді толықтыра түсу мақсатымен мына анықтамаларды беруді жөн көрдік:

1-анықтама. *Қарастырып отырған тілдің, сан жағынан алғанда, шексіз сөйлемдерін талдауға арналған шекті ережелер жиынтығын — талдау моделі деп атайды.*

Модельге қатысты ақпараттар түрі — модельге *енетін ақпарат* және *модельден шығатын* ақпарат болып екіге ажыратылады.

Синтаксистік аналитикалық *модельдерге енетін ақпарат* мәтін түрінде болады да, ал, одан *шығатын ақпарат* — мәтін сөйлемдерінің синтаксистік құрылымы жайлы мәлімет болып табылады.

Семантикалық аналитикалық модельдерге *енетін ақпарат*, жоғарыда аталғандай, мәтін түрінде болады да, ал одан *шығатын ақпарат* арнайы семантикалық тілде жазылған мәтін бойындағы сөйлемдердің мағыналары жайлы ақпарат болып табылады.

2-анықтама. *Шекті ережелер негізінде шексіз көп дұрыс сөйлемдер құрастыратын лингвистикалық модельді жинақтау (синтездеу) моделі деп атайды.*

Синтаксистік синтетикалық модельдерде бастапқы *енетін ақпарат* сөйлемдердің құрылымы жайлы мәліметтерге сай келеді де, ал одан *шығатын ақпарат* — осы тілдің дұрыс құрылған сөйлемдері болып табылады.

Семантикалық синтетикалық модельдерге *енетін ақпарат* арнайы семантикалық тілде жазылған сөйлемнің мағыналық пішіні, ал модельден *шығатын ақпарат* — бастапқы сөйлемдерге синонимдес болып келетін табиғи тілдегі сөйлемдер.

Талдау (анализ) және жинақтау (синтез) модельдерінен басқа, осы аталған модельдер аралығының ортасы болып саналатын *туындатушы* (порождающий) модель туралы да сөз етуге болады.

3-анықтама. *Тілдік белгілердің (символдардың) әліпбиімен бірге сан жағынан шекті туындату ережелерінің негізінде осы тілдің шексіз көп дұрыс сөйлемдерін құрумен қатар, оларға тиісті құрылымдық сипаттамаларды сәйкестендіре алатын қабілеті бар модельді туындатушы (порождающий) модельдер деп атайды.*

Аналитикалық және синтетикалық модельдердің түсінігімен байланысты модельдердің аса бір маңызды ұғымдары: *керіқайтқыштық* (обратимость) және *кері модельдер* ұғымы. Осы ұғымға қатысты жоғарыда (2.3.3 санды тақырыпшада) бастапқы және қорытынды нысандар жайлы айтылғандарға ұқсас келетін **"кері модельдердің"** анықтамасына өзгеше түрғыда тоқталайық.

4-анықтама. *Егер M_1 моделіне енетін ақпарат (нысан) M_2 моделі үшін одан шығатын ақпаратпен сәйкес келсе және, керісінше, M_1 моделінен шығатын ақпарат M_2 моделі үшін енетін ақпаратпен бірдей болатын болса, онда M_1 моделі M_2 моделіне қарағанда "кері модель" деп аталады.*

Кейде зерттеушілер синтетикалық және аналитикалық модельдерді бір-біріне кері модель деп санайды. Егер бұл тұжырымды негіз етсек, екі түрлі (аналитикалық және синтетикалық) модельдерді құрудың қажеттігі болмай қалар еді. Яғни модельдің *керіқайтқыштық* сипаты бойынша аналитикалық түрі мен синтетикалық модельді құруға (және керісінше) болады деген сөз.

Жоғарыда аталған модельдердің математикалық пішінінің баяндалуына қарай, оларды *есептеу* (исчисление) және *алгоритмдер моделі* деп те екіге бөліп қарастырады. Олардың мазмұндық айырмашылықтары: *есептеу моделі – рұқсат ету* жүйесіне қатысты, ал *алгоритмдер – бұйрықтар* (командалар) жүйесіне қатысты. *Рұқсат ету жүйесі* арқылы шекті аппарат көмегімен кейбір жиынның нысандарын толығымен модельге енгізуге болады. Мұндай нысандар саны шексіз болуы да мүмкін. Мысал ретінде қайсыбір тілге қатысты сөйлемдер санын алуға болады.

Рұқсат ету жүйесі сан жағынан аса көп, тіпті шексіз бірліктерді қамтуына байланысты, оны тілді зерттеуге аса ыңғайлы деп санайды.

Енді *алгоритм* ұғымының түсінігіне келсек, ол қалаған нысанды бөліп алу не құрастыру үшін аса қажет рет-ретімен берілетін "бұйрықтардың" (командалардың) тізбегі деп түсіну қажет. Мысалы, берілген мәтіндегі сөздердің әліпбилі жиілік сөздігін құру алгоритмін амалдар топтамалары арқылы іске асатынын сызба арқылы көрсетуге болады. Сөздер санын *"i"* деп белгілесек, $i=1,2,3,...,n$ болады. Яғни мұндай алгоритмнің амалдар топтамаларын мынадай бұйрықтар арқылы орындауға болады:

1) рет-ретімен сөзқолданыстарды мәтін бойынан бөліп ал;

2) соңғы бөлініп алынған сөзқолданысты алдыңғы жазылған сөзқолданыспен салыстыр – кодтық белгілерінің айырымын тап;

3) егер сөзқолданыс кодтарының айырымы нөлге тең болса, сөздің жиілік көрсеткішін "1" санына арттыр;

4) айырым *нөль* санынан аз не көп болса, соңғы бөлінген сөзді сөздік бойына әліпби ретімен орналастыр және жиілік көрсеткішін ***"1"*** санына тең етіп ал.

Міне, жоғарыда аталған мұндай "бұйрықтар" реті мәтін бойындағы сөзқолданыстардың ең соңғысына дейін қайталанып отырады. Мұндай қайталауларды бағдарлама жазу тілінде "цикл" деп атайды. Осындай реттегі жұмыс барысын сызба-топтама (блок-схема) арқылы көрсетіп, оны алгоритм деп атайды. Алгоритмдегі бұйрықтарды компьютерге түсінікті жасанды тілге көшіріліп жазылған түрі – программа (бағдарлама) деп аталады. Кез келген модельдің қызметі алгоритмде орын алған бұйрықтар түріне келтіріліп, сосын ғана оның компьютерлік бағдарламасы жазылады. Себебі қазіргі компьютерлер (өзірше) тек қана жасанды тілде жазылған бұйрық тілін ғана түсіне алады.

Енді жоғарыда сөз болған зерттеу моделіне қатысты *аналитикалық, синтетикалық* және *туындату модельдерімен* бірге *есептеу* мен *алгоритм* ұғымдары арасындағы байланыстарды қарастырайық.

5-анықтама. *Модельдерді құру жолында қандай ережелер түрлері пайдалануына қарай, оларды ықтималдық (статистикалық) және детерминдік (құрылымдық) модельдер деп екіге ажыратып атайды. Ал аралас түрдегі модельдер құрылымды-статистикалық модельдер деп аталады.*

Табиғи тілдің заңдылықтарын аса көп ережені қажет ететін *детерминді (құрылымды) модельдерден* гөрі, ереже саны шектелген *ықтималды модель* арқылы зерттеу ыңғайлырақ деп саналады. *Детерминді модель* арқылы іске асатын дәлдік *ықтималдық моделімен* әрекет етудегі статистикалық ережелердің аздығымен және өзінің қарапайымдылығымен ұтады деуге болады.

Детерминді модельдердің ішіндегі аса құндылары деп фонология мен морфология салаларында пайдаланылатын **"бинарлы дифференциалдық құрылым"** атты модельді, синтаксис саласындағы **тікелей құрастырушы, "трансформациондық"** және **"аппликативті"** модельдерді, ал семантика саласындағы **"семантикалық көбейткіштер"** атты модельді атауға болады.

Лингвистикалық модельдер ішіндегі әрі теориялық та, әрі қолданбалық та жағынан маңызы зор *танымдық модельдерін* атауға болады. Бұл типті модельдер, әсіресе, машиналық аударма, шет тілдерін оқыту және т.б. мәселелердің шешімін табуда көп қолданыс тауып жүр.

Сонымен қатар лингвистикалық модельдерді *парадигматикалық* және *синтагматикалық модельдер* деп те топтауға болады. Бұл модельдердің маңыздылығы олардың тіл қызметіндегі жүйелілік қатынастарды модельдеуге мүмкіндік тудыруында деп санайды. Мысалы, *парадигматикалық модельдер* негізінде тілді жүйе ретінде қарастыра отырып, бір тұрақтылық жағдайдан екінші тұрақтылыққа ауыса алатын *динамикалық модельдерді* құруға мүмкіншілік туындайды. Сол сияқты, тілді салыстырмалы-тарихи әдіспен зерттеу ісін компьютер арқылы жүргізу де осындай *парадигматикалық модельдерді* қолдануды қажет етеді.

2.4 Салалық тілдердің функционалды-коммуникативтік моделі

2.4.1 Модельдеуге қатысты мәтін элементтерін саралау

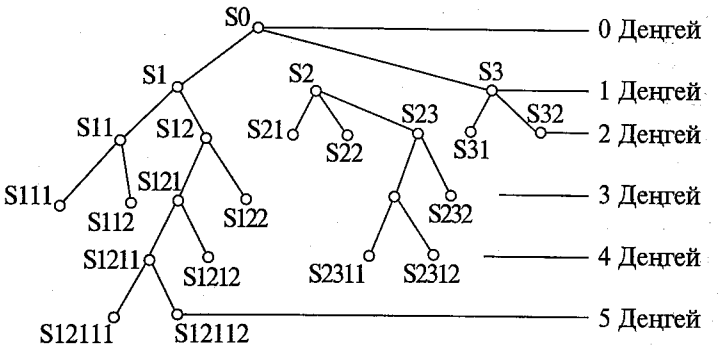
Салалық тілдердің функционалды-коммуникативтік моделін құрудан бұрын ол тілдерді бас жиын ретінде қарастырып, жалпы құрылымын анықтау қажеттігіне сай, мұқият түрдегі дайындық жұмыстарын жүргізудің маңызы зор. Әсіресе, салалық тілдердің машиналық аударма бағытындағы инженерлік-лингвистикалық модельдер үшін аса қажет.

Мәтін түріндегі бас жиынның құрылымы тек зерттеу бірліктерді мұқият түрде сұрыптау арқылы ғана белгілі болады. Яғни қайсыбір белгінің негізінде тілдік бірліктердің иерархиялық таралуы зерттеу нышанының ірілі-кішілі топтарға бөлініп, бара-бара олар аса ұсақ базистік бірлікке дейін бөлшектенуінен байқалады. Міне, осыларды дұрыс жолмен саралау және жинақтай білу модельдеуге қатысты жиынның (мәтіннің) әрбір элементінің номенклатуралық құрылымын ұйымдастыру жұмысына қатысты.

Кез келген тілдік жүйені бір-бірімен өзара тығыз байланыста жатқан элементтердің жиынтығы деп қарастыруға болатыны белгілі. Осындай әрбір жиынды өзінше бір, бөлектенген S_j жүйе ретінде және оларды жалпы сипаттағы S жүйенің (супержүйенің) бөлігі деп, яғни $S_j \subset S$ қарастыруға болады (мұндағы $\subset$ – логикалық белгі: S_j жүйе S жүйенің құрамына енеді деген мағынада қолданылады).

S_j мен S жүйелерінің аралық қарым-қатынасы иерархиялық принциппен құрылады. Яғни S_j-дің S жүйесіне тәуелділігі оларды құрастырушы элементтердің құрылымдағы орнына және функционалды-коммуникативтік бағыт-бағдарына қарай көрініс табады. Осыдан барып әр S жүйенің қасиеттеріне қарай, кішігірім $S_1 - S_2 - S_3$ жүйелеріне сатылап бөліктеу ең кіші базистік элемент алынғанға дейін орын алады. Сонымен бірге S жүйені бөліктеудің әрбір амалдары бөлектенген кішігірім жүйелерді туындатады (S_1; S_2; S_3). Мұндай жағдай S метажүйесінің қайсыбір тарамдану көрінісін бейнелеуге мүмкіндік жасайды. Әрбір кішігірім жүйелер (S_1; S_2; S_3) өз тарапы бойынша әр деңгейде тарамданады. Мысалы, S_1: S_{11}, S_{12}; S_2: S_{21},

S_{22}; S_3; S_{31}, S_{32} және т.т. [Лоскутов, 1972, 8-9-бб.]. Осы айтылғандар *1-суреттегі* сызбада көрініс тапты.



S_0							0	деңгей
S_1	S_2	S_3					1	деңгей
S_{11}	S_{12}	S_{21}	S_{22}	S_{23}	S_{31}	S_{32}	2	деңгей
S_{111}	S_{112}	S_{0121}	S_{122}	S_{231}	S_{232}		3	деңгей
S_{1211}		S_{1212}		S_{2311}	S_{2312}		4	деңгей
S_{12111}		S_{12112}					5	деңгей

Негізгі *S* жүйені кішігірім салалық жүйелерге бөліктеу мен олардың таралу жолдары мына белгілер арқылы іске асады:

1) функционалды-коммуникативті;

2) құрылымдық-жасалу;

3) жүйелік элементтердің құрамы негізінде.

Функционалды-коммуникативтік белгілер бойынша бөліктеу жүйенің айқын қызметі мен коммуникативті бағытына сай жүргізіледі.

Құрылымдық-жасалу белгілері арқылы бөліктеу сол жүйеге тән иерархиялық құрылым ерекшеліктерін ескеру негізінде іске асады. Ал жүйені бөліктеудің соңғы үшінші түрі, яғни элементтердің құрамы бойынша бөліктеу элементтердің қызметі кезіндегі олардың құрылымдарын ескеру арқасында орындалады [Лоскутов, 1972, 8-9-бб.].

Бас жиынның құрылымдық моделі зерттеушінің жеке тәжірибесі мен жинақталған тілдік бірліктердің мәліметтері негізінде құрылуы мүмкін. Тізімделген бірліктердің сұрыпталған функционалды-коммуникативтік белгілеріне қарай топталған және жүйеленген пішінде ақпараттық-іздеу жүйесі құрылады. Осылайша жинақталған мәліметтер зерттеу жиынының *номенклатуралық қоры* деп аталады. Зерттеу нысанының мұндай қоры алгоритмдеуден бұрын орындалатын амал-әрекетке қатысты болуымен қатар, мәтіннің ықтималдық және логикалық құрылымына математикалық статистика әдістері арқылы талдау жүргізуге мүмкіндік туғызады. Әсіресе мұндай талдау ықтималдық түрдегі машиналық аударма мақсатындағы салалық тілдің моделін құруда аса маңызды.

Функционалды-коммуникативтік талдаудың және салалық тілдердің номенклатуралық ұйымдасуы жайындағы жинақталған мәліметтер әрі қарайғы зерттеулерде (алгоритмдеуден тыс) аса қажетті. Әсіресе олар салалық тілдің ерекшеліктерін тануға қатысты дәстүрлі және машиналық автоматтанған сөздіктің сөзтізбесін түзуде ұтымды деп саналады.

Модельдеуге тиісті салалық тіл ретінде, мысалы, *әкімшілік-басқару тіл* (ӘБТ) аясын алуға болады. Ал, ӘБТ-ны модельдеу дегеніміз, сол тілдік жүйенің

элементтерінің құрылымдық ерекшеліктерінің қатынастық бағыты негізінде әкімшілік-басқару тілін (ӘБТ) қайта тудыру деген сөз.

Модельдеуге қажетті ӘБТ өзінің функционалды-коммуникативтік белгісі және таратылатын ақпараттың сипатына қарай, негізгі үш функционалды-коммуникативтік категорияға бөлінеді:

- жоспарлау құжаттары;
- басқару құжаттары;
- ақпараттық есеп беру құжаттары.

Жоспарлау құжаттарының негізгі коммуникативтік функциясы – жағдаятқа баға беру, әрекеттің түпкі мақсатын анықтау, басқару ұйымының, оның бөлімшелерінің, орындаушылардың жұмыс жоспарын құру.

Сонымен, салалық тілді модельдеудегі жоспарлау құжатына қатыстылар: бағалау (estimates), әрекеттің түпкі ойы (concepts), жоспарлар (plans) және т.б.

Басқару құжаттарының негізгі коммуникативтік функциясына жататындар: басқару ұйымының шешімін және алдағы мақсаттарды орындаушының назарына дейін жеткізу. Барлық салалық тілді модельдеу құжаттары (СТМҚ) ішінен ұйымдық-басқару құжаттары (ҰБҚ) ең негізгі және аса маңыздысы болып саналады. Оған жататындар: бұйрықтар, жарлықтар, нұсқаулар және т.б.

Ақпараттық есеп беру құжаттарының негізгі коммуникативтік функциясы – орындалған жұмыстар жайлы, өзінің қызмет жағдайы және т.б. жайлы тиісті орындарды (инстанцияларды) хабардар етіп тұру. Құжаттардың бұл категориясы өзінің әр түрлігімен ерекшеленеді. Модельдеуге қатысты салалық тілде мұндай құжаттарға жататындар: әрқилы мәліметтер (summaries), хабарламалар (reports), шолулар (surveys), зерттеулер (studies), әрекеттер мен оқиғаларды талдаулар (evaluations) және т.б.

Бұл аталған ӘБТ-ның негізгі категориялары іштей тағы да құжаттардың үлгілеріне байланысты түр-түрге ажыратылады. Ал соңғылар өз алдына құжат шығарушы бөлімшелердің түрлеріне, басқарушы ұйымның қызметіне және т.т. байланысты әрі қарай да *1-ші суретте* көрсетілгендей топ-топқа бөлінуі мүмкін.

ӘБТ жайлы арнайы атап кететін жайт, бұл тіл саласының құрылымдық жағынан өзіне тән үйлесімділік пен қисындылық және бірізділік пен қатал тәртіппен байқалатын функционалды-коммуникативтік бағыт-бағдарының болуы.

Тәжірибелік байқауға қатысты бірліктер функционалды-коммуникативтік категория негізінде айырым табады. Ол бірліктер іштей тағы да типтерге, ал ол типтер базистік бірліктерге ажыратылады. Осы аталған ӘБТ *1-кестеде* көрсетілді.

1-кесте

Категориясы	Типі	Түрі (базистік бірлік)	Барлығы	Жинағы
Жоспарлау құжаттары	Түпкі ой Бағалау Жоспар	1 9 10	20	
Басқару құжаттары	Тез орындалатын Күнделікті Заңгерлік	17 8 5	30	
Ақпараттық есеп беру құжаттары	Тез орындалатын Күнделікті	67 8	75	
3	8	125	125	125

Кестеде бірліктердің бір орынға жинақталып берілуі иерархиялық принципке негізделген. Модельдеу жиынтығында барлығы үш түрлі функционалды-коммуникативтік категория қамтылған. Әрбір категория 8 типтік құжаттарға және 125 әртүрлі құжаттарға бөлінген [Нелюбин, 1975].

2.4.2 Тілдік бірліктерді кодтау және салалық тілдің формалдауға жататын элементтері

Мәтінді лингвостатистикалық талдау мен мәтіннің инженерлік-лингвистикалық моделін құруда оның сыртқы тұрпатынан және безендірілуінен байқалатын ерекшеліктерін де ескеру қажет болады. Тілдік бірліктер мен мәтіндерге тән ерекшеліктерді жүйелі түрде **кодтау ісі** (шартты белгімен ауыстыру) құжаттарды ЭЕМ арқылы өңдеу жұмыстарында маңызды болып келеді. Мысалы, ұйымдық-басқару құжаттары стиліндегі ортақтық пен ерекшеліктер ақпаратты сипаттаудағы сыртқы тұрпаттың өзінен-ақ байқалады.

Кез келген табиғи тілдің ақпаратты тарату мен оны қабылдау жылдамдығы адам баласының есте сақтау мүмкіндігіне сәйкестендірілген. Психологтардың зерттеулеріне қарағанда, адамның көру мен естуінің ең үлкен мүмкіндігі 50 бит/сек жылдамдық шамасымен шектеледі екен [Кондратов, 1978, 53-б.]. Бұл жердегі "бит" – ақпараттың екілік жүйедегі өлшем бірлігі. Мысалы, 1 секунд ішінде адам көлемі тек 50 әріптен (дыбыстан) аспайтын ақпаратты ғана қабылдай алады екен. Сондықтан айтылым кезіндегі ақпаратты қабылдаушы адам аз уақыт ішінде барынша мол мәлімет қабылдап алу үшін оның сыртқы материалдық пішін саны да барынша аз болған жөн.

Тілдік ақпаратты қабылдаудың ең ұтымды жолы оның материалдық пішінін статистикалық сипаттағы таңдалып алынған шартты белгілермен (кодтармен) ауыстыру [Борисов, 1972, 51-б.]. Ол үшін мәтінді модельдеу кезінде оның бөлек элементтерінің тұрпаттық жақтары ескерілуі қажет.

Тілді толығымен формалдаудың мүмкін еместігі адамның ойлау процесінің логикалық жағын толық формалдаудың қиындығынан деп түсіндіріледі [Глушков, 1964].

Математика тілімен айтқанда, оған тілдік жүйенің ашық жүйеге қатыстылығы да көп кедергі тудырады. Солай бола тұра, сол ашық жүйенің ішінен-ақ, белгілі түсінікте, өзінше шектелген және формалдауға жататын элементтері бар кішігірім жүйелер қатарын кездестіруге болады. Мұндай жүйе ретінде шағын тілдік стильге жататын мәтіндер қарастырылады. Бұл мәтіндерді, құрылымы мен шекті элементтері тұрғысынан қарағанда, формалды лингвостильдік жүйе деп ұйғаруға болады.

Осындай шағын тілдік стильге тән формалдауға жататын элементтерді функционалды-коммуникативтік талдау негізінде бөліп алып, оны көптеген инженерлік және лингвистикалық мәселелерді анықтау қажеттігіне пайдалануға болады.

Шағын тілдік стильдер жүйесіне тән формалдауға жататын элементтер – *құрылымдық, шартты, еркін* және *семантикалық элементтер* деп төрт топқа бөлінеді [Нелюбин, 1983, 25-42-бб.]. Келесі тақырыпшаларда осы аталған топтарға қысқаша тоқталамыз.

2.4.3 Құрылымдық формалдау элементтері

Құрылымдық формалдау – мәтіндегі оқиға мәліметтерін (деректерін) баяндаудың тәсілі. Мұндай формалдау тәсілі шағын стильдің номенклатуралық құрамын топтау негіздемесінен бастап, сол мәтіннің сыртқы композициялық құрылымдарына дейін қоса қамтиды.

Құрылымдық формалдау белгілі дәрежеде тілдік салалардың бәріне бірдей тән болғанымен, іс қағаздары мен қызмет бабында жазылған хат мәтіндерінің жиынында көбірек көрініс табады.

Мәтіннің сыртқы композициялық құрылымы болып табылатын рубрикалау – мәтінді құрылымдық формалдау тәсілдерінің бірі. Рубрикалау формуляр мен құжат мәтінін құрамдас бөліктерге бөлуге сәйкес келеді. Ол бөліктердің мәтін ішіндегі графикалық ажыратылу түрлері мынадай: *тақырыптар, тақырыпшалар, тармақтар, тармақшалар, параграфтар, параграфшалар, абзацтар, нөмірлеу* мен *пунктуациялар.*

Тақырып пен тақырыпшалар құрылымдық формалдаудың элементтеріне жататындығымен қоса, олар мәтін (құжат) мазмұны жайлы шағын және ұтымды түрде құжат бөлігінің тақырыбын және негізгі ой түйінін бере алады. Сондықтан ол нақты ақпаратты қамтитын хабарламаның маңызды бірлігі болып саналады. Сол сияқты, тыныс белгілері де формалдау элементіне жататындығымен қатар, оларда семантика-синтаксистік және функционалды-коммуникативтік мән-мағына жатыр. Мысалы, бір тілден екінші тілге компьютер арқылы аударма жасаудағы шағын стильдік саланы модельдеу жұмыстарында және т.б. тыныс белгілердің маңызы ерекше.

Мәтіндік құжаттың жалпы құрылымын және функционалды-коммуникативтік бағытын қарастырғанда оның сыртқы пішінін ескеру қажет. Іс құжатының пішіні оны безендіру мен мазмұнын анықтайтын элементтердің жиынынан тұрады. Құжат түрі сол элементтердің құрамы, көлемі және олардың орналасу реті арасындағы қатынастық байланысты білдіретін мәліметтер арқылы ажыратылады [Рахманин, 1973, 242-б.]. Құжат формасының бөлектелген элементтері – *реквизит*, ал олардың жиынтығы мен белгілері *құжаттың формуляры* деп аталады [Нелюбин, 1983, 27-б.].

Модельдеуге тиісті мәтіндерді талдай келе, біз олардың композициялық қисындылығын, мәлімет берілісінің логикалық үйлесімділігі мен шағын және ұтымды түрде баяндалуын байқауға болады. *Құжат формулярының* бірізді және тек осы функционалды-коммуникативтік типтегі құжатқа ғана тән үлгімен айрықша құрамды *реквизит* элементтерінен тұратынын байқатады. Аталған құжат құрылымы оны оңай қабылдауға мүмкіндік тудырады және компьютерді дұрыс пайдалануды жеңілдетеді.

Формулярлардың реквизит құрылымы, яғни оның саны, мағынасы мен реті кейбір мәтіндерде біркелкі болмай, көптеген жағдайда құжаттың фунционалды-коммуникативтік бағыты мен номенклатуралық қатыстығына байланысты болады. Ал кейбір *реквизиттер құрылымы* барлық құжаттарға ортақ болып келуі мүмкін.

Құрылымдық жағынан алғанда *формуляр* көбіне мынадай *реквизиттерден* тұрады:

1) құжат түрлерінің атауларына сәйкес жалпы тақырып (штамп);

2) құжаттың нөмірі, датасы (күні, айы, жылы), құжат иесі, яғни құжат шығарушы басқару органы және басылым орны;

3) араб, рим цифрларымен және әріптерімен бөлінген мәтін параграфтары мен тармақтары;

4) жауапты мекеме аттары мен оған қатысты адам аттары аталған қорытынды және жауапты адамның лауазымы мен қолы.

Құрылымдық формалдау элементтерінің *формуляры* мәтін түрінде немесе графиктік және типтік (трафареттік) пішінде болуы мүмкін.

Мәтінді *рубрикациялау* мен *құрылымдық формалдауға* қатысты "абзацтың" да маңызы зор. Сондықтан құжаттық мәтіндерді абзацқа бөлу бірізділікті және аса дәлдікті қажет етеді. Мысалы, абзац (азат жол) дәл үш бос орыннан кейін барып жазылуы қажет болса, құжат бойы сол қалып сақталуы керек. Сол сияқты әрбір тақырып, тақырыпша, тармақ, тармақша алдын ала келісілген "бос орындардан" (пробел) кейін жазылуы шарт.

Құжат ішіндегі абзацтар бір не бірнеше сөйлемдерді қамтуы мүмкін. Ал *пунктуация* тек рубрикалау мен құрылымдық формалдау элементтері ғана болып қоймай сонымен бірге нақты семантика-синтаксистік және функционалды-коммуникативтік мағыналы да жүк атқара алады. Сол сияқты, келісі қарастырылатын **"шартты формалдау элементтерінің"** де орнына жүре алады.

2.4.4 Шартты формалдау элементтері

Шартты формалдау тәсілі тыныс белгілерін айрықша түрде қолдану, баяндаудың қалыптасқан сыртқы пішінін қолдану және шартты белгілерді, символдар мен реквизиттерді пайдалану арқылы іске асады.

Әдетте мәтіндерді (ақпараттарды) автоматты түрде талдау мен жинақтауда тыныс белгілер жайлы мәселе ескеріле бермейді [Малаховский, 1971]. Бірақ біртілдік және екітілдік жағдайларындағы ақпаратты автоматты түрде өңдеу мақсатында тыныс белгілердің беретін мәліметтері аса көп.

Мәтін ішіндегі тыныс белгілер тек графикалық жүйенің бір бөлігі ғана емес, ол әліпбилік белгінің (кодтың), грамматикалық және лексикалық мүмкіншіліктерімен айқындала қоймайтын мағыналық қатынастарын реттей алады, яғни айтылымның бағыт-бағдарын белгілеп, қажетті семантикалық арақатынасқа нұсқап отырады. Сонымен, мәтін бойындағы тыныс белгілер әрі грамматикалық, әрі семантикалық, әрі стильдік те қызметтерді атқара алады деуге болады [Кобрина, Малаховский, 1961].

Осы айтылғандармен қатар, тыныс белгілер шартты формалдау элементтерінің де орнына жүреді. Ақпарат мағынасы дұрыс ашылуы үшін оны таратушы мен қабылдаушы шартты түрде қолданылған тыныс белгілердің өзгеше кодын өзара түсінісе білулері керек.

Компьютерлік аударма және ақпаратты автоматты өңдеу, оны іздеу тәжірибелерінде модельдеуге қажетті мәтін бойында (орыс және ағылшын тілдері бойынша) кездесетін он үш тыныс белгінің тек оны ғана пайдаланылып жүр. Мысалы, оған жататындар: нүкте, нүкте-үтір, үтір, қос нүкте, қысқа сызықша, жақшалар, сыңар тырнақша, ұзын сызықша, қисық жақша және сұрақ белгісі (сирек жағдайда). Шартты белгілер мен символдар барлық құжаттарда қолданылады. Олар бір сөз, бір әріп, әріптер мен цифрлар тіркесі, бірнеше цифрлар тіркесі және топографиялық белгілер арқылы көрініс табуы мүмкін.

Бір сөзбен берілген шартты белгілер латын әліпбиіндегі әріптер арқылы, яғни фонетикалық (радиотелефондық) әліпбидің сөз-символының көрінісінде болады. Мысалы: *ALFA – A, TANGO – T, WHISKEY – W, XRAY – X.*

Бір әріппен берілген шартты белгілер, мысалы, *Z* – дирижабльді, *V* – ұшақты білдіреді.

Шартты формалдаудың әріппен берілуінің ерекшелігі олардың көпшілігінің алдын ала тіркелген ондаған мағынаға ие болуында. Мысалы, *"Z"* әрі *"дирижабль"*, әрі *"Гринвич уақыты"* және т.т., барлығы 29 мағына; *"M"* – "тез-тез емес" (асықпау) белгісі, ол барлығы 136 мағына және т.б.

Бірақ бір әріптік шартты белгінің қолдануы формулярдың реквизиттеріндегі, біріншіден, алатын орын тәртібінің қатаң болмауымен, екіншіден, белгілі бір құжат түрінде тұрақты бір ғана мағынада берілуімен ерекшеленеді. Сондықтан мұндай шартты формалдау элементтері машиналық аударма жұмыстарында көп мағыналықты шешуде көп қолданылады. Ал модельденетін мәтіндерде бірнеше цифрлардың тіркесімен берілген шартты белгілер уақыт пен датаны көрсету үшін қолданылатын. Мысалы, "дата - уақыт" тобы *6* цифрдан тұрады, оның екеуі – айдың күнін, ортаңғы екеуі – уақытты, ал соңғы екеуі – минут шамасын білдіреді. Егер осы аталғандар бір орынды санмен берілсе, олардың әрбіреуінің алдына "нөл" қойылады. Мысалы, шартты белгілер тобы 010509 түрінде берілсе, оны айдың *1* күні *5* сағат *9* минут деп ұғыну қажет.

Шартты белгілер әріп, цифр және шартты таңба арқылы берілуі метеорологиялық, яғни ауа райын болжау құжаттарында кездеседі. Мысалы, *2Ø ⊖A 25⊘ ⊕* – *two thousand scattered, pilot reports ceiling two five thousand overcast.*

Соңғы жағдайдағы шартты таңбалар арқылы беріліс құжат мәтінін 90-95% қысқартуға дейін мүмкіндік тудырады.

Компьютердің көмегімен мәтінді статистикалық тәсілмен зерттеу жағдайында "нөлдік" белгі, яғни "бос орын" (пробел) белгісі де кейде шартты формалдау элементі ретінде алынады.

Бұл аталған шартты белгілер графика түріндегі модельдеу құжаттарында кең қолданыс табады. Мұндай құжаттардағы әріппен жазылған мәтін өзінің негізгі рөлін жоғалтып, оның орнына қосымша мағынаға ие болатын шартты символдар негізіндегі ақпарат болып табылады.

2.4.5 Еркін формалдау элементтері

Еркін формалдау элементтерінің түсінігінде мәтін бойындағы **тілдік бірліктерді** (ТБ), сол сияқты толық сөйлемдер мен кейбір ұғымдарды алдын ала келісілген цифрлар мен әріптер арқылы және олардың аралас түрлерімен ауыстырып жазу процесінің ұғымы жатыр.

Қарастырып отырған мәтіндегі еркін формалдау элементтеріне жататындар: арнайы және қызметтік сигналдар мен көмекші кодтар (фразалар коды), шақыру сигналы, парольдер мен жауаптар және арнайы қабылданған шифрлар мен кодтар.

Арнайы сигналдар дегеніміз, ол кейбір команданың, бұйрықтың және хабар берудің шартты белгілерінің қысқарған түрі. Бұл аталғандар қайсыбір сан, сөз және цифрлар мен әріптердің аралас түрінің көрінісі болуы да мүмкін. Арнайы сигналдарды қолдану ақпаратты тез тарату мен қабылдау процесін жеделдетеді. Әсіресе, ол байланыстың техникалық тораптарында орын алады. Мысалы, олар радиотораптар аралығында пайда болатын өзара кедергілерді азайтады.

Арнайы сигналдар қауіп-қатерді хабарландыру үшін қолданылса, ол хабарландыру сигналдары деп, ал қызметаралық келіссөздер кезінде қолданылса –

қызметтік сигналдар деп екіге ажыратылады. Бұл сигналдар арқылы жүргізілетін байланыстың әрі жылдам, әрі сапалы болатындығы дәлелденіп отыр.

Қызметтік сигналдар түріне қызметтік таңба, қызметтік сөз және қызметтік фразаларды жатқызуға болады. Олар бірәріптік, екіәріптік және үшәріптік топтарды құрап, белгілі бір мағынаға ие сөзден немесе сөз тіркестерден тұруы мүмкін. Бұл аталғандар кейбір жиі қолданылатын тапсырыстардың, сұраныстардың, нұсқау мен мәлімдемелердің қысылған пішінін байланыс торабы арқылы тарату кезінде кең қолданылады. Мысалы, *K* – қабылдауға көшем, *A* – байланыс соңы, *ROGER* – сізді түсіндім, *OUT* – байланыс соңы.

Қызметтік сигналдарға үшәріптік топтан тұратын фразалар кодын (қызметтік код) да жатқызуға болады. Үшәріптік топтар латын әліпбиінің *Q* немесе *Z* әріптерінен басталып жазылатынын ескеру керек.

Фразалар коды формалдаудың өте тиімді құралы деуге болады. Себебі олар арқылы толық бейнелі сөйлемдерді ауыстыруға мүмкіндік туады. Мысалы: *ZUA – уақыт сигналы қазір беріледі, QRU – менің сізге басқа айтатын тапсырмам (хабарым) жоқ.*

Фразалар кодына цифрларды не қызметтік белгілерді қосып жазу арқылы олардың семантикалық мағыналары нақтылануы немесе кеңеюі мүмкін. Мысалы, *ZNB* фразалық код *"пороль"* мағынасын білдіреді. Егер оған 2 саны қосылып, ZNB2 болып жазылса, оның мағынасы былайша кеңейеді: *ZNB2 – соңғы хабарламаның поролі.*

Ал фразалық код сұраулы сөйлемді білдірсе, оның алдына *"INT"* қызметтік белгі қойылады. Мысалы, *INT QRU – маған радиограмма бар ма?* Фразалар коды, негізінен, қайсыбір ауызша түрдегі тапсырыстар, сұраныстар, нұсқаулар мен мәлімдемелер, бұйрықтар және т.б. тарату үшін қолданылады.

Модельденуге тиісті шағын тілдік аяны (подъязык), мысалы, құжат мәтінін арнайы шифрлар мен кодтар арқылы толық формалдауға мүмкіндік туады.

Кодталған мәтін (закодированный текст) – арнайы кодтар көмегімен жазылған мәтін. Код дегеніміз шартты түрдегі қысқарған белгілердің жинағы кодтау тәсілі бойынша мәтін бойындағы әрбір сөз, сөз тіркес, сөйлем шартты түрдегі қысқарған сөздермен, символдармен және олардың тіркесімдерімен ауыстырылады.

Кодталған мәтін, көбінде, дипломатиялық, әскери және сауда қызметтерінің мәліметтерін тарату ісінде қолданылады.

Шифрланған мәтін – арнайы шифр арқылы жазылған мәтін. Бұл жердегі "шифр" дегеніміз құпия хаттарды жазуға арналған шартты белгілер жүйесі, яғни өзінше бір құпияжазба (тайнопись) түрі. Шифрлау кезінде әрбір бөлек әріп алдын ала келісілген басқа бір әріппен, цифрлармен немесе белгімен ауыстырылып жазылады.

Мәтіннің кодтау мен шифрлау электронды-есептеу машиналарына (компьютерлеріне) қажетті ақпаратты енгізу мен оларды өңдеуді барынша жеңілдетеді. Олар өзінше бір формалды тіл болуынан ақпаратты барынша "қысылған" түрде жазуға мүмкіндік жасайды.

Сонымен, еркін формалдау элементтерін қолдану, яғни қызметтік сигналдарды, кодтарды, шифрларды, шақыру сигналдарды, шифрларды, парольдер мен жауаптарды және т.б. қолдану қажетті деген ақпаратты тарату мен қабылдау кезінде уақыт мөлшерін барынша қысқартуға, мәтінді белгілі дәрежеде "қысуға" мүмкіндік туғызады.

2.4.6 Семантикалық формалдау элементтері

Семантикалық формалдау элементтерді аббревиацияның (қысқарту тәсілі) барлық тәсілдерінің мәтін ішінде кең түрде көрініс табуы деп түсінген жөн. Модельденуші шағын салалық тілдерде аббревиация тетігі белсенді сипатта болады. Мысалы, кейбір мәтіндерде қысқартылып берілген тілдік бірліктер барлық сөзқолданыстың 60 пайызына дейін жетеді, ал ондай аббревиациялар сол мәтін сөздігінің 16,3% құрайтыны анықталып отыр.

Тілдік бірліктерді кең түрде қысқартып беру тәсілі тарату ақпаратының көлемін бірнеше рет және мәтіннің графикалық пішінін 95 пайызға дейін "қысуға" мүмкіндік туғызады. Әрине, қысқартуды да жөнімен, яғни мәтін мазмұнын ашуға теріс әсері болмауын есепке алу керек. Байланыс тораптары инженерлерінің зерттеу нәтижелері бойынша ақпарат мазмұнына зиянсыз "қысу" мөлшері 50% аспауы қажет деп саналады [Анисимов, 1959, 12-б.]. Мұндай "қысудың" пайыздық мөлшері әр тілге қатысты айырым табатыны белгілі. Мыселен, Л.Бриллюэннің пайымдауынша, ағылшын тілінде мәтінді "зиянсыз қысу" үш есеге дейін өседі екен [Бриллюэн, 1969, 49-50-бб.].

Аббревиация тәсілінің қажеттігі мен мүмкіндігі көптеген экстралингвистикалық факторларға байланысты болып келеді. Бұл жердегі сыртқы және ішкі факторлардың өзара қисындылығы қоғам өмірі мен адамдардың негізгі қарым-қатынас құралы – "тілге" тығыз байланысты. Сондықтан аббревиация кездейсоқтық құбылыс және қайсыбір адамдардың қалауынан қабылданып тұрған жоқ, ол қарым-қатынастың объективті және заңды қажеттілігінен туындайды. Тілдің даму заңдылығы оның қоғамдық функциясымен анықталады және ол тіл құрылымының заңдылығымен тығыз байланысты болып келеді.

Ең бір негізгі экстралингвистикалық фактор ретінде модельденуші шағын тілдік стильге тән анықтық, нақтылық, дәлдік, шағындық және динамикалық сипатына тікелей қатысты **уақытты үнемдеу факторы** болып саналады.

Әртүрлі селективті ақпаратқа ие болатын әртүрлі ақпараттық-семантикалық жүк артатын және сөйлеу желісінде ақпаратты тарату мүмкіндігі бар мәтін элементтерінің түрлері (әріп, сөз және т.б.) адам баласының ақпаратты қабылдау мен өңдеу процесіне қатысы бар ми жұмысының үздіксіз сипатынан гөрі, бөліктелген (квантталған) ақпарат ағымына ыңғайлы болып келеді.

Қысқарту тәсілі арқылы ақпарат аз мөлшердегі белгілер негізінде таратылады және мәтін белгілеріне қарағанда аббревиация белгілерінің "көтеретін" ақпараттық жүгі барынша "ауыр" деуге болады [Медникова, Шеляховская, 1971, 107-б.].

Сонымен, аббревиация тәсілі ақпарат тарату нәтижелерінің тиімді жолы болып саналады. Себебі ол тәсіл арқылы ақпарат тарату жолындағы неше түрлі "кедергілер" жойылады және ақпарат тұрпатының өзі бірізділік пен ұтымдылыққа мүмкіндік туғызады.

Аббревиация нәтижесінде түпнұсқадағы тілдік бірлік (ТБ) құрылымдық жағынан өзгеріске ұшыраған және барынша қысқарған жаңаша пішіндегі тілдік бірлікке (ТБ) ауысады.

Е.П.Волошинның айтуынша, аббревиацияның негізінде ақпаратты таратудың лексика-грамматикалық комплекстерді қамтитын мүмкіндіктері ішінен жинақы бірлік тұтастығы – сөз тіркестерді қысқартып беру жатыр [Волошин, 1967].

Қысқарған сөздер, өздерінің тілдегі атқаратын қызметіне қарай: *графикалық* және *лексикалық* деп екіге ажыратылады.

Графикалық қысқарту қарым-қатынастың тек жазба түрінде қолданылады және олар тілдесу кезінде өзінің бұрынғы толық түрінде айтылады. Мысалы, *ҚР — Қазақстан Республикасы.*

Лексикалық қысқартулар тілде әріптердің әліпбиде аталуымен (АҚШ, БҒМ) немесе әдеттегі сөздерді айту нормасымен дыбысталады [Нелюбин, 1976, 21-24-бб.].

Қысқартудың аталған екі түрі де лингвистикалық бірліктерге (ЛБ) қойылатын талаптарға толығымен жауап береді. А.И.Смирницкийдің айтуынша, тілде бөліп алуға болатын бірлік тіл бірлігі болуы үшін, біріншіден, оның сыртқы (дыбыстық) түрі және сол сияқты мағыналық немесе эмоционалдық мазмұны болуы керек. Екіншіден, тек коммуникация кезінде ғана пайда болатын емес, сонымен бірге әр уақытта тілде қолданыс таба алатын дайын тілдік шығарма ретінде де іске асуы қажет. Тек осындай екі шартты бірдей қанағаттандыратын бірлікті тілдік бірлік деп атауға болады [Смирницкий, 1956, 13-б.]. Лексикалық қысқартудан туындайтын бірлік осы аталған екі қасиетке де ие.

Графикалық қысқартуға тән бірліктер жайлы мынаны айтуға болады. Ондай бірліктер қатарына жататын жазба коммуникация түріндегі **толық тілдік бірліктердің** көпқасырлық қолдануының өзі, олардың мәтін бірлігі екендігіне дау туғызбайды. Себебі графикалық қысқарту адамның тек көріп қана қабылдауына арналған [Борисов, 1972, 117-б.].

Инженерлік лингвистика тұрғысынан алғанда жоғарыда аталған қысқартулардың екі түрі де ТБ (тілдік бірлік) деп қарастырылады. Қысқартулар сөзформа (сөзтұлға) ретінде "адам — машина — адам" жүйесінде басқа тілдік бірліктермен бірдей дәрежеде қызмет атқарады. Олар көзбен көріп қабылдау мен екітілдік жағдаяттарда аса ыңғайлы түрде іске асып жүр. Басқа сөзбен айтқанда, қысқартылған пішіндегі тілдік бірліктер коммуникация талабын қанағаттандыра алатын толыққанды тілдесу бірлігі деуге болады, яғни дербес мағынаға ие болуымен қатар коммуникация кезінде де өзіне тән дыбысталу (және мәтіндерде графикалық) пішінде қолданыс табады.

Аббревиация процесінде туындайтын трансформациялық ережелердің әсері негізінде қысқартулар — морфологиялық және синтаксистік аббревиациялар деп екіге ажыратылады. Синтаксистік аббревиация түрі тілді "қысуға" (компрессия) қатысты тіл бірліктерінің құрылымын өзгерту (трансформация) ережелерінің әсерінен пайда болады. Олар тілдік құрылымның барлық деңгейін, яғни жоғарғы сөйлем деңгейінен бастап, төменгі графемдік (фонемдік) деңгейді қоса қамтып жатады [Нелюбин, 1969].

Біз, негізінен, тек қысқаша түрде морфологиялық аббревиация мәселесіне ғана тоқталамыз.

Редукцияланатын элементтеріне қарай, графемді-морфологиялық деңгейде пайда болатын негізгі қысқарту әдістері төмендегідей айырым табады:

1) Бір сөзді "қысу", оның бастапқы әрпімен берілсе, ол **моно-инициалды аббревиация** деп аталады. Мысалы: *Москва=M; amplitude=A; west=W; east=E.*

2) Егер көп компонентті термин немесе құрама сөздің компоненттерін құрайтын бірнеше сөз, олардың бастапқы әріптерімен "қысылса" және олар бір тілдік бірлікке (ТБ) дейін қысқаратын болса — **поли-инициалды аббревиация** деп аталады. Мысалы: *Америка Құрама Штаты — АҚШ, Біріккен Ұлттар Ұйымы — БҰҰ, Ғылым академиясы — ҒА, баспасөз — БС, loudspeaker* (дыбыс күшейткіш) — *LS (ГГ).*

3) Сөздің қысқарып жазылуы бастапқы және соңғы әріптерінен тұрып, ал кейбір жағдайда ол сөздің ішкі әріптерінің де қатысы болса, ондай қысқарту түрі **рамалық аббревиация** деп аталады. Мысалы: *team (команда) — TM; day(день) — DY; train(поезд) — TN; letter(письмо) — LTR; traffic(движение) — TFC.*

4) Сөздің қысқарып жазылуы оның бастапқы бөлігінен (бірнеше бастапқы әріптерінен) тұрса, оны **моно-морфемді аббревиация** деген терминмен атайды. Мысалы: *equalizer – EQ; movement – MOV; motor – MOT; Tuesday – TUE.*

5) Көпкомпонентті термин құрайтын бірнеше сөздер олардың алғашқы буындарына дейін қысқарып, бір тілдік бірлікке (ТБ) дейін қысылса, олар **полиморфемді аббревиация** деп аталады. Мысалы: *medical laboratory (медицинская лаборатория) – MEDLAB.*

6) Жоғарыда аталған қысқарту түрлері араласып кездессе, **құрастырылған аббревиация** деп аталады.

Сонымен, формалдау элементтеріне қатысты жоғарыда қарастырылған мәселелер негізінде мынандай қорытынды жасауға болады. **Мәтін реквизиттерінің** көпшілігі: *клишеленген конструкциялар (құрылыстар), жұмсалу (айтылу), үлгілер (стандартные обороты) немесе сөйлеу өрнектері (словесные формулы), шартты белгілермен мен қысқартулар.* Осы аталғандардың негізінде модельденуші шағын саланың (подъязык) мәтінін белгілі дәрежеде *шартты*, *еркін* және *семантикалық элементтері* арқылы формалдау жолдары іске асады.

Инженерлік психология саласының мәліметтері бойынша, мұндай стандартталынған айтылымдар ақпаратты қабылдауды 8-10 рет тездетеді екен [Малитков, Сольский, 1972, 94-б.]. Себебі мұндай құжаттардың құрамында көптеп кездесетін шартты белгілер мен қысқартулар шартты код сипатына келтіріліп, әрі қысылған, әрі қысқа мазмұнға әкеледі. Сондықтан мұндай **формуляр реквизиттері** мен мәтіндерді бірізділікке келтіру мен ақпаратты таратудың уақытын ғана үнемдеп қоймай, сонымен бірге оны қабылдауда да аса ыңғайлы. Сонымен аталған тәсіл модельденуші "шағын тілге" қатысты ақпараттық амалдарға автоматтандыру құралдарын қолдануға барынша мүмкіндік тудырады.

Үшінші тарау

ТАБИҒИ ТІЛДІҢ СИНТАКСИСІ МЕН СЕМАНТИКАСЫН МОДЕЛЬДЕУ

3.1 Табиғи тілдің синтаксисін модельдеу

Қолданбалы лингвистика саласындағы әрқилы мәселелердің шешімін табуда, әсіресе, жазба не сөйлеу мәтіндерін автоматты түрде өңдеу кезінің ең бір маңызды этапы – **сөйлемдерге синтаксистік талдау жүргізу**. Ал мұндай талдаудың негізгі мақсаты – сөйлем мен оның бөліктерінің синтаксистік құрылымын танып-білу, сөйлем құрамына енетін сөзтұлғалар тізбегінің сол тілдің алдын ала белгілі грамматикалық ережелеріне сай келуі мен құрылымдық шарттардың орын алу мүмкіндіктерін анықтау. Егер осындай алдын ала белгілі шарттар қанағаттандырылса, талдауға алынған сөйлем осы грамматикаға қатысты "дұрыс сөйлем" деп саналып, оның құрылымы анық түрде сипатталады. Сөйлемдердің синтаксистік құрылымын сипаттау әдістері арқылы табиғи тілді зерттеу – **синтаксистік құрылымдар теориясының** [Буторов, 1996, 142-160-бб.] негізгі проблемаларының бірі.

Аталған теорияны сөз етуден бұрын, синтаксистік талдауды қажет ететін мәтінді автоматты өңдеудің кейбір тұстарына тоқталайық. Оған жататындар: толық сөйлемнің немесе оның бөлігінің, мысалы, сөйлем құрамындағы есімді сөз тіркестердің синтаксистік құрылымын анықтау.

Бір тілден екінші бір тілдерге машиналық аудару жүйелерінің дамыған түрлерінің барлығында да автоматтанған синтаксистік талдау жүргізіледі. Мұндағы синтаксистік талдау дегеніміз, біріншіден, сөйлемнің семантикалық түсінігін алу қажеттігі болса, екіншісі – белгілі бір тілдегі аударуға тиісті сөйлемнің синтаксистік құрылымын "аударма грамматикасы" ретінде аталатын екінші тілдік жүйедегі сөйлемдер құрылымына ауыстыру.

Ақпараттық іздеу жүйесінде мәтінді (немесе құжатты) талдау кезінде кейде толық емес немесе толыққа жуық түрдегі синтаксистік талдау қолданылады. Мұндай жағдайда сөйлем бөліктерінің құрылымы, негізінен, сөйлем құрамындағы есімді сөз тіркестері танылады. Оның себебі, мұндай жүйедегі мәтінді автоматты өңдеудің мақсаты оны индекстеу, яғни мәтіндегі сөздер мен сөз тіркестері арқылы берілетін құжаттардағы атаулар ұғымы мен терминдік сөз тіркестерді автоматты түрде тану (ажырата білу) болып табылады. Осындай талдау жағдайында тек қана есімді сөз тіркестердің синтаксистік құрылымын сипаттайтын "грамматика" бірден-бір қажет.

Электронды-есептеу машиналары (ЭЕМ) мен адам аралық қатынасты табиғи тілде іске асыруға әрекеттенетін "жасанды интеллект" жасау мәселесіндегі мәтінді тану (түсіну) жүйесінде табиғи тілдегі сөйлемдерге синтаксистік талдауға қатысты зерттеу жұмысы, әдетте, жеке бір бөлек топтама (блок) құрайды.

Сол сияқты, сөйлеу тілін тану мен түсінуге қатысты автоматты жүйе де өзінше бөлек семантика-синтаксистік топтаманы құрайды. Ондай топтама айтылымның грамматикалық және семантикалық дұрыстығын тексеру бағытын ұстаным етеді.

Бұл бағыт бойынша, мағыналары ақырына дейін анықталмаған сегменттік сөз тізбектерін тану мақсатымен қосымша жорамал (гипотеза) жасалады.

Сонымен бірге, зерттеушілердің синтаксистік құрылым теориясына қызығушылығы ішкі лингвистикалық қажеттілік тұрғысынан да туындайды. Мысалы, сөйлем құрылымының дәл және формалды берілісі анықталмайынша, табиғи тілге тән синтаксистік параметрлердің кең спектріне статистикалық зерттеулер жүргізу мүмкін емес.

Қазіргі кезде сөйлемнің синтаксистік құрылымының көрінісін білдіретін негізгі үш түрлі әдіс кең қолданыс тауып жүр.

Оларға жататындар:

1) тікелей құрастырушылар құрылымы немесе құрастырушылар таралуы;

2) қатынастар құрылымы немесе қатынастар таралуы;

3) аталған екі жағдайды бірге қарастыратын құрылымдар.

3.1.1 Құрастырушылардың құрылымы

Құрастырушылар моделінде немесе құрастырушылар жүйесінде сөйлем сөзтұлғалардың реттелген сызықтық тізбегі (П) ретінде қарастырылады. Дәлірек айтсақ, тізбек құраушы тілдің ең кіші бірліктері – морфемалар, сөзтұлғалар, ал кейбір жағдайда одан да үлкен болуы мүмкін, мысалы, тұрақты сөз тіркестері.

Осындай реттелген сызықтық тізбек (П) құрайтын кесінділердің жиыны құрастырушылардан тұратын "С" жүйесі деп аталады, олар төмендегі шарттарды қанағаттандырулары керек:

1) "П" сөйлемінің өзі де "С" жиынына оның элементі ретінде енеді, яғни "П" реттелген сызықтық тізбегі (П) құрастырушылар жүйесі "С"-ның бір бөлігі.

2) "П" тізбегіндегі барлық сөзтұлғалар, яғни бірлік ретінде алынған кесінділер "С" жүйесін құрайтын жиынға қатысады.

3) "С" жүйесі жиынындағы кез келген екі кесінді бір-бірімен қиылыспайды немесе олар бір-біріне енеді.

Сонымен бірге, толық сөйлем құрайтын "П" және бөлек сөзтұлғалар "С" жиыны құрастырушыларының жүйесіне енеді.

"С" жүйесінің жиынын құрайтын сөйлем кесінділері – құрастырушылар деп, ал сөйлемнің өзі "П" – жиынның толық құрастырушысы деп аталады. Бөлек сөзтұлғаларға сәйкес келмейтін және толық емес жиын құрастырушылары сөз тіркес ретінде түсінік табады.

Осындай түсініктегі құрастырушылар жүйесі **сөйлемінің синтаксистік құрылымының моделі** ретінде қарастырылады. Аталған модельде элементарлық синтаксистік бірлік ретінде бөлек сөзтұлғалармен бірге сөз тіркестер де көрініс табады. Ал сөз тіркестердің құрылымдық қатынасы ретінде бір сөз тіркесінің екінші сөз тіркесіне тәуелді болуы (қатыстылығы) ескеріледі. Сөз тіркестер сөйлем құрамындағы синтаксистік қатынаста тұруына байланысты оларда "жекелік қиылысу" болмауы керек.

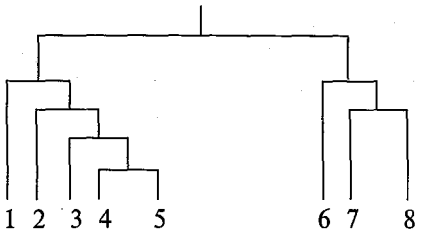
"С" жиындағы бір құрастырушы екінші құрастырушының құрамына тікелей ену мүмкіндігін анықтауға болады. Ал "С" жиынына енетін құрастырушылардың бір-бірімен қатынастары граф сызбасы (дерево зависимостей) немесе "тәуелділіктер тармақтары" арқылы көрініс табады. Ондай көріністе граф төбешігі ретінде ең кіші бірлік ретінде – сөзтұлғалар алынады. Мұндай "тәуелділіктер тармақтарын"

құрастырушылар таралуы немесе "НС" таралуы деп те атауға болады. "С" жиынын тәуелділіктер тармақтары арқылы графикалық түрде жазу көрнектілік үшін аса қажет. Құрастырушылар таралуы "П" сөйлемнің құрастырушылар құрылымы деп те аталды.

Сөйлемнің "НС" таралуына мысал ретінде В.Д.Буторовтың мақаласында мынадай сөйлем қарастырылған:

1) *"Для (1) представления (2) синтаксической (3) структуры (4) предложения (5) используется (6) структура (7) составляющих (8)"* [Буторов, 1996, 145-б.].

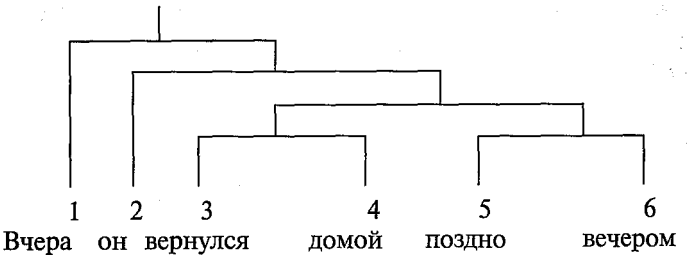
Енді осы сөйлемнің графикалық жазылуы былайша көрініс тапқан:



Осы көріністі, яғни сөйлемнің құрастырушы құрылымын жақша арқылы да жазуға болады. Мұндай тәсілде әрбір элементтен тәуелділіктің дәрежесіне қарай, төменде көрсеткендей, жақша ішіне енгізіледі:

((1(2(3(4 5)))) (6(7 8)))

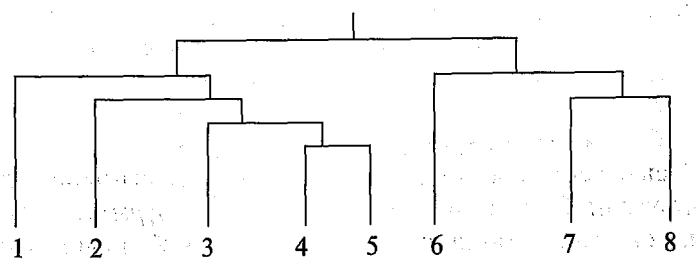
Сол сияқты, аталған мақалада мынандай екінші сөйлем де қарастырылған – *"Вчера он вернулся домой поздно вечером"* [Буторов, 1996, 145-б.]:



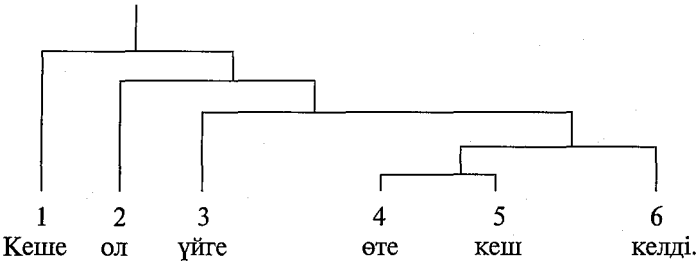
(1((2((3 4) (5 6)))

Енді осы сөйлемнің қазақша аудармасының құрылымын граф көрінісі мен тәуелділіктер сызбасы арқылы қарастырайық:

1) *Сөйлемнің (1) синтаксистік (2) құрылымын (3) анықтау (4) үшін (5) құрастырушылардың (6) құрылымын (7) пайдаланамыз (8).*

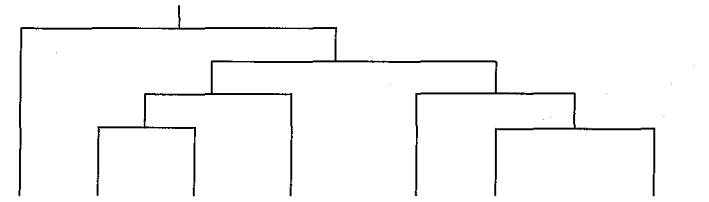


2) Кеше (1) ол (2) үйге (3) өте (4) кеш (5) келді (6).

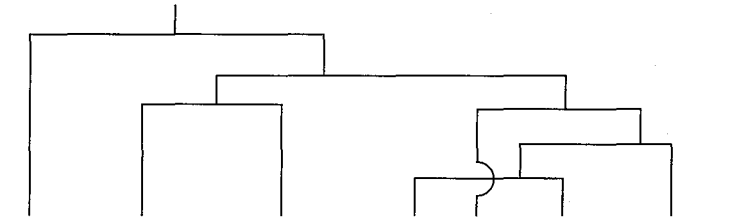


Бірақ құрастырушылар құрылымы жоғарыдағы көрініс түрінде жеңіл салына бермейді. Осы айтылғанның дәлелі ретінде мақала авторы А.С.Пушкин өлеңдерінен мынадай жолдарды келтірген [Буторов, 1996, 146-б.]:

Он из Германии туманной привез учености плоды.



Осы сөйлемнің мағынасы (түсінігі) екі түрлі болуы мүмкін. Біріншісінде – *"тұманды Германиядан"* (жоғарыда көрініс тапқан сызба), екіншісі – *"ғылымдық жемістің тұмандылығынан" ол Германиядан әкеледі.*
Екінші түсінік бойынша:



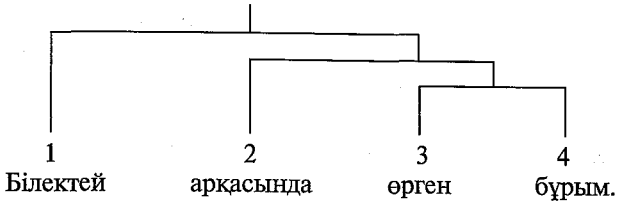
Шынында да, Германияның ауа райын ағылшын жерінің ауа райымен салыстырсақ, оны өне бойы тұман түсіп тұратын жер деуге келмейді. Сондықтан біріншіге қарағанда, екінші граф көрінісі шындыққа сай келеді.

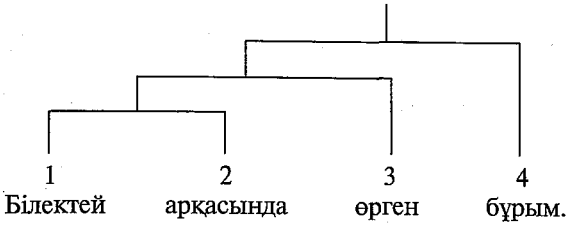
Сол сияқты, Абай өлеңдерінен мына жолдарды қарастырайық:

"Білектей(1) арқасында(2) өрген(3) бұрым(4).
Шолпысы сылдыр қағып жүрсе ақырын".
Көмшат бөрік, ақ тымақ, қара қасты,
Сұлу қыздың көріп пе ең мұндай түрін.

"Білектей" сын есім сөзі *"арқа"* зат есім сөзінің сынына қатысты деген түсінік те болуы мүмкін. Сондықтан *"білектей"* сөзінің *"бұрымға"* қатыстылығы – ең дұрыс шешім. Соңғы айтылғанға сай келетін бірінші сызба шындыққа жақын тұр.

1)





Құрастырушылардың, кем дегенде біреуі, кесінді түрінде болмаса, мұндай құрылым үзікті құрастырушылар құрылымы деп аталады. Үзікті құрастырушылар құрылымы көркем әдебиет стиліндегі сөйлемдерде айтылымның ритмикалық құрылымы үшін қажет болады.

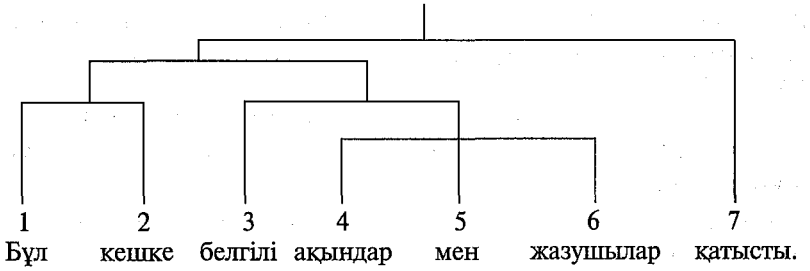
Сөйлем құрастырушыларын талдаудың қиыншылығына қарамай, құрастырушылар моделі немесе құрастырушылар жүйесі сөйлем құрылымын грамматикалық тұрғыдан да, семантикалық тұрғыдан да дұрыс сипаттайды деуге болады. Мұндай дұрыстық сипаттағы құрастырушы сөзтұлғалар "омонимді сөз" болып, сөйлем мағынасын ажырату қажеттігі кезінде де және "омонимді сөйлем" мағынасын ажырату кезінде де байқалады.

Келесі қарастыратын сөйлем:

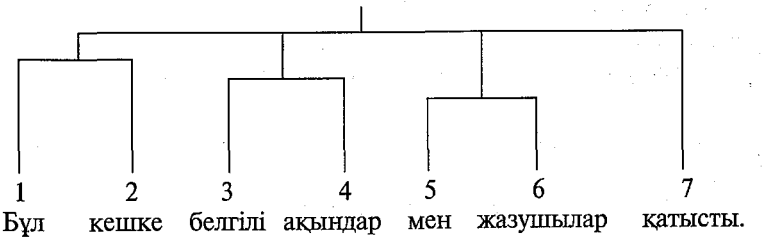
Бұл(1) кешке(2) белгілі(3) ақындар(4) мен(5) жазушылар(6) қатысты(7).

Осы сөйлемді екі түрлі мағыналық құрылымда қарастыруға болады:

1) Бірінші құрылымда ***белгілі*** сын есім сөзін тек ***ақын*** зат есім сөзге қатынасты десек:



2) Екінші құрылым жүйесін *"белгілі ақындар"* және *"кейбір жазушылар"* (*белгілі жазушы* болуы шарт емес) деп түсінуге болады.



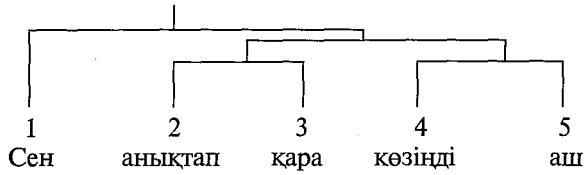
Құрастырушылар жүйесі моделінің синтаксистік деңгейде омонимдік сөйлемдерді ажырату мүмкіндігі оның негізгі ұтымдылығы деп саналады. Мысалы, мынадай екі сөйлемді қарастырайық:

1) Сен(1) анықтап(2) қара(3), көзіңді(4) аш(5).

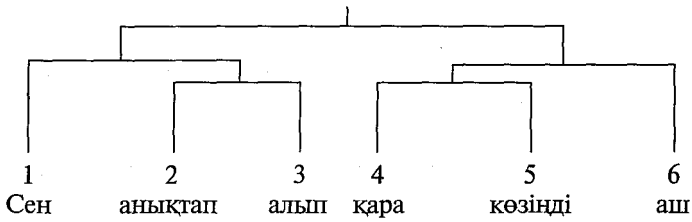
2) Сен(1) анықтап(2) алып(3), қара(4) көзіңді(5) аш(6).

Бұл сөйлемдердегі ***қара*** сөзі омонимдік қатар құрайды. Олардың құрастырушылар таралымы төмендегідей көрініс табады:

1) Бірінші құрылымда ***қара*** сөзінің етістікке қатыстылығы көрініс тапты.



2) Екінші құрылымда ***қара*** сөзінің сын есімге қатыстылығы көрсетілді.



Егер құрастырушылардың құрылымына олардың түрі жайлы қосымша ақпарат енгізетін болсақ, құрастырушылар моделінің (жүйесінің) лингвистикалық мәні арта түседі.

Сол сияқты, әрбір құрастырушы сөзге не сөз тіркеске, оның топтық аражігіне қарай, шартты таңба (метка) қойылған болса, ондай құрастырушыларды таңбаланған (размеченное) құрылымдық құрастырушылар деп атайды (размеченная структура составляющих).

Мәселен, құрастырушылардың туындау грамматикасында (Гладкий, Мельчуг, 1969; Фитиалов, 1984) туындайтын сөйлемдерге таңбаланған таралу құрастырушылар таңбасы қойылған.

Құрастырушылар грамматикасы сөйлем мағынасының дұрыстығын емес, ол оның туындау синтаксисінің дұрыстығын модельдейді.

Мақалада құрастырушылардың мынандай грамматикалық типтері таңбаланған [Буторов, 1996, 149-150-бб.]:

S – сөйлем,

$O_в$ – обстоятельство времени (мезгіл пысықтауышы),

$O_м$ – обстоятельство места (мекен пысықтауышы),

$N_{вин}$ – группы существительного в винительном падеже (табыс септігіндегі зат есім тобы),

$A_{им}$ – группа прилагательного в именительном падеже (атау септігіндегі сын есім сөздер тобы),

$P_{вин}$ – предлог с управл. на вин. падеже,

$Р_{тв}$ - предлог с управл. в твор. падеже,
$П_{им}$ - предлог с управл. в им. падеже,
$П_{вин}$ - предлог с управл. в вин. падеже,
$С_{им}$ – существ. в им. падеже,
$С_{вин}$ – существ. в вин. падеже,
$С_{тв}$ – существ. в тв. падеже,
V – глагол в личной форме,
Н – наречие.

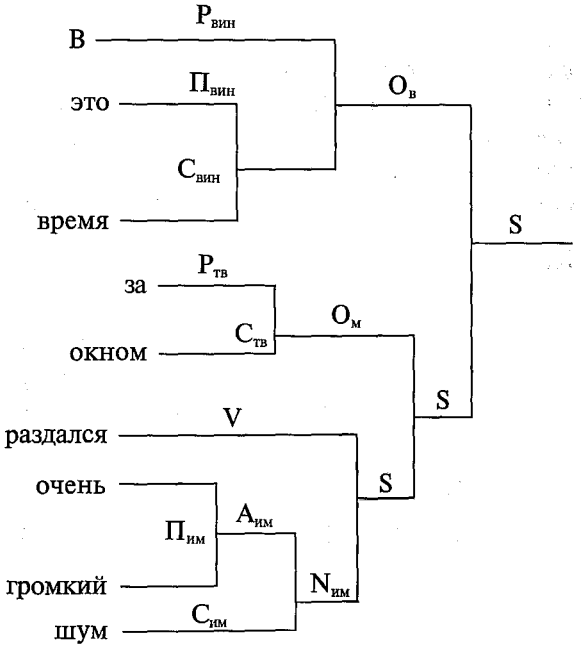
Аталған мақалада орыс тіліндегі мына сөйлем қарастырылған: *"В это время за окном раздался очень громкий шум"*.

Бұл сөйлемді таңбалау үшін құрастырушы грамматиканың төмендегідей ережесі пайдаланылған [Фитиалов, 1984, 91-б.]:

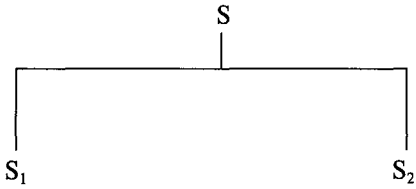
$S \rightarrow О_в\ S$
$S \rightarrow О_м\ S$
$S \rightarrow V N_{им}$
$О_в \rightarrow Р_{вин}\ N_{вин}$
$О_м \rightarrow Р_{тв}\ С_{тв}$
$N_{вин} \rightarrow П_{вин}\ С_{вин}$
$С_{им} \rightarrow$ шум
$С_{вин} \rightarrow$ время
$С_{тв} \rightarrow$ окном
$V \rightarrow$ раздался
$Н \rightarrow$ очень
$N_{им} \rightarrow А_{им}\ С_{им.}$
$А_{им} \rightarrow НП_{им}$
$Р_{вин} \rightarrow$ в
$Р_{тв} \rightarrow$ за
$П_{им} \rightarrow$ громкий
$П_{вин} \rightarrow$ это

Мұндай грамматикалық ережені қолдану арқылы сөйлемдердің синтаксистік құрылымының таңбаланған құрастырушылар жүйесін (моделін) сипаттауға болады.

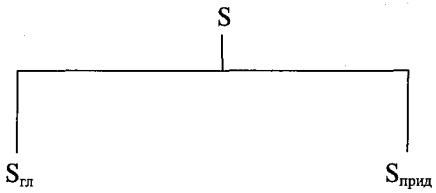
Сөйлемдерді олардың құрастырушылары бойынша талдаудың кейбір ережелерін айқындау қазіргі кезде дәстүрге айналуда. Автор олардың кейбір жақтарын ғана қарастырады.



1) Салалас құрмалас сөйлемдер S_1 және S_2 құрылымдары граф арқылы былайша көрініс табады:



2) Бағыныңқы құрмалас сөйлемдер $S_{гл.}$ және $S_{прид.}$ мына құрылымда бірігеді:



$S_{гл.}$ – басыңқы сөйлем, $S_{прид.}$ – бағыныңқы сөйлем.

Егер бағыныңқы сөйлем түгелдей басыңқы сөйлемге қатысты болмай, оның құрамындағы бір негізгі сөзге (слово-хозяин) ғана тәуелді болса, онда бағыныңқы сөйлем сол сөзбен бірігіп, бөлек құрастырушы құрайды.

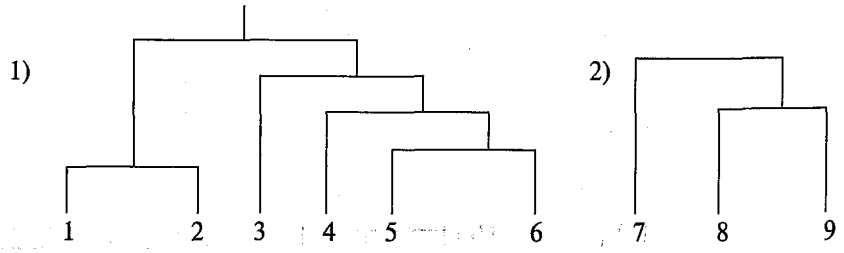
Мыс. ***Он(1) знал(2), что(3) был(4) прав(5)***



Зат есім сөздің оны анықтайтын анықтауыш сөзімен толымсыз немесе толымды түрде қиысып, тіркесуі мүмкін. Бұл тіркестің синтаксистік жақындығын тек семантикалық критерийі арқылы анықтауға болады. Осыған байланысты мақалада мынандай екі сөйлем қарастырылған:

1) Цифровые(1) вычислители(2) со(3) встроенным(4) устройством(5) контроля(6);

2) Текстовая(1) база(2) данных(3).



3.1.2 Тәуелділіктердің құрылымы

Сөйлем ішіндегі сөзтұлғалар арасындағы сызықтық тәуелділік бинарлы қатынастық тәуелділікке жатады. Кез келген сөйлем құрамындағы екі сөзформадан тұратын сөзтіркес грамматикалық және мағыналық жағынан қарастырғанда біркелкі бола бермейді. Мұндай тәуелділіктерді: қабысу, жанасу, қиысу, меңгеру; меңгерілу деп жалпылама сипаттағы дәстүрлі түрде қарастыруға болады. Сөйлемдер ішінде кездесетін сөзтұлғалар арасындағы мұндай тәуелділіктер жиыны сол сөйлемнің синтаксистік құрылымы жайлы мол көлемді ақпарат бере алады.

Тәжірибе жүзіндегі ғылыми ізденістерде мұндай тәуелділіктердің жиыны, яғни сөзтұлғалардың сөйлемге кірігу жиынында, көптеген жағдайда, **"тәуелділіктер тармақтары"** түріндегі графтық көрініс құрайтыны анықталды. Бұл жердегі **"тәуелділіктер тармақтары"** деген сөз тіркес орыс тіліндегі "дерево зависимостей" деген терминнің орнына қолданылып тұр. Ал **сөзтұлғаның сөйлемдегі кірігуі** – "вхождение словоформ в предложение" дегеннің баламасы.

Анықтама: "П" сөйлемінің таралым тәуелділігі деп, сөйлемге сөзтұлғалардың кірігу жиынын құрайтын ең ақырғы графты атайды, егер:

1) сөйлем құрамында басқа сөзтұлғаларға тәуелсіз, ең болмағанда, бір сөзтұлға табылса. Бұл сөзтұлға таралудың негізгі тамыры (корень дерева) немесе таралудың төбешігі (вершина дерева) деп аталады;

2) "П" сөйлемінің таралу төбешігінде орын алған сөзтұлғадан басқа кез келген сөзтұлға басқа сөзтұлғалардың тек біреуіне ғана тәуелді болса;

3) графта тұйық жолдар орын алмаса.

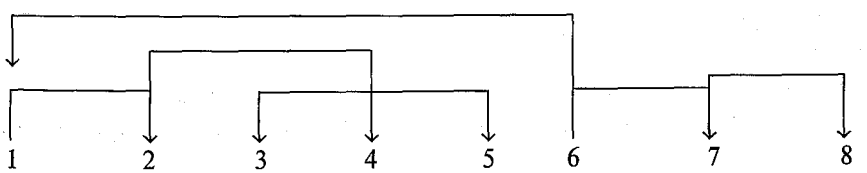
Әдетте, таралу тәуелділігінде көрініс тапқан "жолдар ізі" сөзтұлғалардың тізбегін құрайды. Бұл тізбектегі әрбір келесі кездесетін сөзтұлға алдыңғысына тәуелді болып келеді. Әрбір таралудың қиылысу торабын оның негізгі төбешігімен (тамырымен) қосатын тек жалғыз ғана "жол" түседі. Егер бұл "жолдың" басы мен аяғы беттесіп жатса, ондай жолды тұйық жол деп атайды.

Егер "П" сөйлеміндегі ***x*** сөзтұлғасынан "жол" ***y*** сөзтұлғасына апарса, онда ***x*** сөзтұлғасы ***y*** сөзтұлғасын өзіне жанама түрде бағындырады (немесе тәуелді етеді) деп атайды. Басқаша айтсақ, ***y*** сөзтұлғасы ***x*** сөзформасына жанама түрде тәуелді болады. Осы жағдайды әдебиетте басқаша терминмен де айтады: ***x*** сөзтұлғасы ***y*** сөзтұлғасын басқарады. Осындай ***x*** сөзтұлғасына жанама сипатта (тәуелді) бағыныңқы сөзтұлғалар жиынын осы ***x*** сөзінің "тәуелді тобы" деп атайды.

Ал ***y*** сөзтұлғасының ***x***-ке тікелей тәуелділігі жанама тәуелділіктің жекелік жағдайына қатысты деп саналады.

Екі сөзтұлғаның біріне-бірінің қатыстылығы (тәуелділігі) былайша белгіленеді: $x>y$.

Қорытынды ретінде, жоғарыда сөз болған, сөйлем ішіндегі сөзтұлғалардың синтаксистік тәуелділігін айқындайтын ғалым В.Д.Буторовтың "Моделирование синтаксиса естественного языка" атты мақаласындағы мына сөйлемді келтірейік [Буторов, 1996, 153-б.]: *Для(1) представления(2) синтаксической(3) структуры(4) предложения(5) используется(6) дерево(7) зависимостей(8).*



3.2 Сөйлем семантикасын модельдеу

3.2.1 Синтаксистік семантиканың дамуына қатысты сыртқы әсерлер

Сөйлем семантикасы зерттеу нысаны болу кезеңі өткен ғасырдың 60-жылдарынан басталады деуге болады. Себебі, одан бұрын, тіл білімінің қолданбалы саласына қатысты бұл тақырып жеке зерттеу аясы болмағаны белгілі.

Сөйлем семантикасының моделін құру жолымен айналысуға үш түрлі жағдаят өз әсерін тигізді деуге болады [Богданов, 1996, 161-200-бб.]. Оның біріншісі – тіл білімінің ғылым ретінде қалыптасуындағы ішкі логикалық даму процесінің әсері десек, екіншісі – сыртқы әсер, яғни экстралингвистикалық әсер, ал үшіншісі – лингвистиканың өзімен жақын жатқан ғылымдар аясымен белгілі дәрежеде байланыста болудың әсері.

Көпшілік зерттеушілердің пайымдауынша, сөйлем құрылымын формалды сипаттаудың қандай жетілген түрі болса да, олар зерттеушіге тілдік қарым-қатынастың табиғатына үңіліп, ішкі сырын жақсы түсінуге жағдай жасай алмайды. Яғни, формалды сипаттау "алмасу ақпаратының" мазмұндық сырын ашуға мүмкіндік туғыза бермейді. Сондықтан зерттеу нысаны ретінде ақпараттың мән-мағынасының өзін алуға қажеттік туындайды. Осы айтылғандар ғалым-зерттеушілердің жоғарыда көрсеткен бірінші әсерін дәлелдейді.

Ал синтаксистік семантиканың дамуына қатысты, екінші сыртқы әсер ретінде, зерттеу аясының "жасанды интеллекті" жасап шығаруға қызығушылығы себеп болды. Бұл жердегі "себеп" деп отырғанымыз – адам баласының ойлау әрекеті мен компьютерлік әрекеттердің Н.Винерден, А.Тьюренгтен т.б. бастама алатын, ғылымдағы ұқсастықтар жайлы идеялар. Келесі бір себепке қатысты мәселе – интеллектуалды автоматтың адаммен қарым-қатынасындағы сөйлеу не жазба түріндегі табиғи тілде түсінісу мүмкіндігін тудыру. Мұндай мүмкіндіктер қазіргі кезде, белгілі дәрежеде, іске асып отыр деуге де болады.

Енді, үшінші әсерге жататындар – тіл біліміне шекаралас және жақын жататын философия, әлеуметтану, психология, информатика және т.б. ғылымдар салаларының әсері. Синтаксистік семантикаға осындай әсердің ең жемістісі деп символдық логика саласы саналады. Қазіргі кезде логика саласының идеялары мен терминдер атаулары синтаксистік семантикада кең түрде орын алуда.

Сөйлем семантикасын модельдеуде, ең алдымен, жоғарыда аталған сөйлемнің мағына-мәні ретінде "нені" қабылдаймыз, деген сауалды анықтап алуымыз қажет. Әрине бұған бірден жауап беру қиын. Бірақ көптеген зерттеушілердің тұжырымдауынша, сөйлем мағынасы аса күрделі және көпаспектілі құрылым. Себебі сөйлем мағынасында аса күрделі түрдегі мына жайттар шоғырланған: болмыстың экстралингвистикалық сипаты, концептуалды құрылым ретіндегі оның адам сана-сезімінде бейнеленуі және әр тілге тән өзіндік ерекшеліктері.

Сөйлем семантикасын модельдеуде осы аталған аспектілердің әрбіреуі негізгі дәрежеге ие. Егер бір себептермен, басқа аспектілер есепке алынбаса, ол қазіргі синтаксистік семантика теориясы бойынша – аспектінің жалғыздық сипатта болуынан.

Сонымен, синтаксистік семантиканың көпаспектілі құрылымының тек біреуі ғана негізгі болып саналады да тек сирек жағдайларда ғана олардың саны екеуге жетуі мүмкін.

Қай аспектінің бастапқы, яғни тірек аспекті ретінде алынуына қарай, сөйлем семантикасын модельдеу үшін төрт түрлі ұстаным негіз болып жүр.

Оның бірінші үшеуі "әлем – ойлау – тіл" үшбұрышы элементтерінің қатыстылығына қарай есепке алынады да, ал төртінші ұстаным – "тіл – адам" прагматикалық қатынасқа негізделеді.

Әрі қарайғы баяндауымызда келтірілген төрт түрлі зерттеу ұстанымы мынандай терминдермен аталады:

1) онтологоцентристік;
2) концептоцентристік;
3) синтактикоцентристік;
4) антропоцентристік.

Осы жолдардың әрбіреуіне қатысты белгілі дәлдікте алынған сипаттауларға қысқаша тоқталайық.

3.2.2 Онтологоцентристік бағыт

Жағдаят ұғымы. Бұл әдіс бойынша сөйлем мағынасының теориясын құруда алғышарт ретінде "онтология" алынады. Яғни бұл жердегі "онтология" дегеніміз тіл арқылы көрініс табатын экстралингвистикалық ақиқат болмыс немесе тілден сырт жатқан болмыс.

Мұндай тілден сырт жатқан болмыстың сөйлем арқылы сипатталатын ең кіші бөлігі ретінде жағдаятты немесе деректі алу келісілген. Сондықтан сөйлем мағынасын зерттеу "жағдаят" ұғымы мен оларды топтау ұғымының түсінігін білуді қажет етеді. Бұлардың әрқайсысына (жағдаят пен топтау) нақты сөйлемге қатысты белгілі бір семантикалық құрылым сәйкестендіріледі. Сөйлем семантикасын модельдеуде "жағдаяттан" сөйлемге өту былайша іске асады. Ең алдымен, ақиқат болмыста жекеленген немесе эмпирикалық (тәжірибелік) "жағдаяттың" болуы мүмкін деп қабылданауы керек. Аталған "жағдаяттардың" бір-бірінен құрылым жағынан немесе кеңістіктегі алатын орнына және заттық нысанның саны мен сипатына қарай ерекшеленетіні ескеріледі. Жекеленген эмпирикалық "жағдаят" нақты айтылым арқылы көрініс табатыны айқын. Осының салдарынан айтылым бөліктері "жағдаят" бөліктеріне сәйкес келуі керек. Сонымен бірге айтылымдарда "жағдаят" бөліктері арасындағы қатынастар да өз орнын алуы қажет.

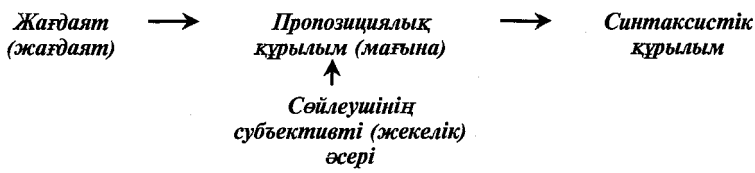
Нақты эмпирикалық жағдаяттарды топтастыра отырып, кластарға жинақтау қажеттігі туындайды. Басқаша айтсақ, абстрактылы "жағдаятқа" саналы түрдегі талдаулар мен қорытындылау қажет болады. Мұндағы абстрактылы "жағдаятқа" айтылымдардан гөрі көбінде толық сөйлемдер түрі сәйкес келеді. Қарапайым түрдегі "жағдаят" қарапайым оқиға немесе "атомдық факт" деп аталады. Тілдік деңгейде мұндай "жағдаяттар" – қарапайым немесе өзекті сөйлем арқылы сипатталады.

Сонымен, онтологиялық модельдеуде әлемнен тілге қарай бетбұрыс ретін мынандай негізгі үш сатымен көрсетуге болады:

(1)		*(2)*		*(3)*
Жағдаят	⟶	***Пропозициялық***	⟶	***Синтаксистік***
(жағдаят)		***құрылым***		***құрылым***
(онтология)		***(мағына)***		***(тілдік беріліс)***

Мұндағы "пропозиция" (П) – сөйлем мен одан туындайтын құрылымдардың модальды және коммуникативті парадигмаларына ортақ болатын семантикалық инвариант (түр). Ең алғашында, "пропозиция" сөйлем ұғымын білдірген. Пропозицияның көлемі сөйлемнің бір бөлігі деуге болады немесе ол сөйлемнің номинативті (атауыш) не семантикалық (мағыналық) аспектісін білдіретін бөлік.

Бұл сатылар ақиқат болмыста орын алатын ойлау мен тіл өрісіне сәйкес келеді. Әсіресе мұндағы 1-ші және 2-ші сатылар зерттеушіге қажеттілеу. 1-саты "жағдаяттың" адам сана-сезімінде бейнеленуіне байланысты болады. Мұндай "жағдаят" сызба түріне көбірек үйлеседі, бірақ "пропозициялық құрылым" (мағына) "жағдаят" құрылымына "изоморфты" түрде ғана қатысы бар. Оның изоморфты болуы үшін жағдаяттың әрбір компонентіне "пропозициялық құрылымның" белгілі бір элементі сәйкес келуі керек. 1-ші сатыға қарағанда, 2-ші саты аса күрделі процестен тұрады, себебі мұнда сөйлеушінің субъективті факторы, яғни сөйлеушінің "даралық" ерекшелігі орын алады. Сонымен бірге, "пропозициялық құрылымға" сөйлеушінің коммуникативтік ниеті мен пропозициялық мазмұнның құндылығына беретін "бағасы" және т.б. орын алады. Тек осы аталғандар ескерілсе ғана, "пропозициялық құрылым" синтаксистік нысан ретінде, яғни сөйлем ретінде іске асады. Осы айтылғанды сызба түрінде былайша көрсетуге болар еді:



Сөйлеушінің пропозицияға өзгеріс енгізетін әсері, пропозициялық және синтаксистік құрылымының толық изоморфты болуына көбінде жол бермейді.

3.2.3 Референция ұғымы

Референция дегеніміз – айтылымның болмыспен байланысы және осындай қатынасты іске асыратын тілдік құралдар. Басқаша айтқанда, **референция ұғымы** біреудің (сөйлеушінің) белгілі бір мәселе туралы мақұлдап немесе сынап білдіретін пікірі деуге болады.

Семантиканы модельдеудегі онтологиялық тәсілде экстралингвистикалық "жағдаят құрылымы" ескерілмеуі мүмкін, бірақ референция тұжырымдары ескерілуі қажет. Референция тек нақты айтылым жағдайында, яғни сөйлеу кезінде ғана мүмкін болады деп саналады. Яғни референция тек коммуникативтік акт кезінде ғана іске асады. Сондықтан ол, белгілі дәрежеде, прагматикамен қатынасы бар. Бұл жердегі сөз болып отырған прагматика – сөйлеушінің болмыстағы қайсыбір белгілерге қатынасы. Ал **сөйлеушінің референциясы** –сөйлеушінің болмысты белгілеу мүмкіндігі. Сондықтан айтылым мазмұнында семантикалық компонентпен бірге прагматикалық компонент те болуы қажет. Бірінші компонент, яғни **семантикалық компонент** "сөйлемнің потенциялды концепті" ретіндегі пропозициямен беріледі. Екінші компонент, яғни **прагматикалық компонент** прагматикалық айнымалылардан тұрады. Оған жататындар: **сөйлеуші, тыңдаушы, сөйлеу актісіндегі орын мен уақыт, жалпы көзқарас, жалпы білім қоры**. Егер айнымалылар нақты константаларымен, яғни тұрақты деректермен алмастырылса **сөйлемнің потенциалды концепті шынайы мағыналы тұйық пропозицияға** ауысады деп пайымдауға болады.

3.3 Концептоцентристік бағыт

Концептоцентристік әдісте әлемнің экстралингвистикалық құрылымын анықтау қарастырылмайды. Сондықтан бұл теория бойынша, бастама ретінде "жағдаяттың" орнына сөйлем негізінде жататын **"концептуалды құрылым"** алынады. Мұндай құрылым экстралингвистикалық деректерсіз-ақ, тікелей тілдік берілістен тұрады.

Концептоцентристік теорияның қазіргі кезде ең көп қолданыс тауып жүрген түрі – **"символдық логика"** ұғымының кейбір түріне негізделетін және негізгі ұстаным ретінде **"пропозиция ұғымы"** алынады. Ал логика саласындағы пропозиция ұғымы концептуалды құрылым ретінде түсіндіріледі. Табиғи тілде мұндай құрылымға жататындар: лексикалық немесе грамматикалық константтар. Пропозиция ұғымын ұстаным ететін **семантиканы модельдеу теориялардың** барлығы да шартты түрде **"пропозициялық бағыттағы теориялар"** деп аталады.

Семантиканы модельдеу ісі логикалық пайымдаулар арқылы да іске асады. Мұндай модельдеудің пропозициялық бағытқа қатысы жоқ.

Пропозициялық бағыт бойынша семантиканы модельдеуде сөздердің септелу қатынасын ескермейтін де және оны басшылыққа алатын да теориялар орын алады. Олардың орыс тіліндегі әдебиеттердегі атаулары: "теории, не учитывающие падежно-ролевых отношений"; "падежно-ролевые теории моделирования семантики предложения". Мұндай теориялар жайлы, яғни пропозициялық және пропозициялық емес бағыттар жайлы мәліметтерді В.В.Богдановтың "Моделирование семантики предложения" атты мақаласынан толығырақ танысуға болады [Богданов, 1996, 168-181-бб.].

3.4 Синтактикоцентристік бағыт

Сөйлем семантикасын талдау ісі жоғарыда аталған әдістерге қатыссыз, сөйлем ішінде орын алатын синтаксистік сызбаға негізделіп қана жүргізілуі мүмкін. Мұндай талдауға қатысты зерттеу жұмысы, алғашында, сөйлемнің құрылымдық сызбасын ажырату мен оларды бөліктеп, жүйеге келтіру мен құрамын анықтау арқылы іске асты.

Бұл әдіс жайлы толық мәліметтермен ғалым Н.Ю.Шведованың 1980 жылы жарық көрген "Русская грамматика" кітабынан танысуға болады. Аталған грамматикада **"құрылымдық сызбаның семантикасы"** және **"сөйлемнің семантикалық құрылымы"** ұғымдары енгізіледі және осы ұғымдардың ұқсас жақтары сөз болады.

Семантикалық құрылымдары ұқсас, бірақ грамматикалық тұрпаты әртүрлі болып келетін барлық сөйлемдер типтері өзара формалды-семантикалық қатынаспен байланыста болатындығы жайлы айтылған. Синтактикоцентристік әдіс онтологияға толығымен қатыссыз болуымен ерекшеленеді.

Кейбір тілшілер лексикалық толығудан тыс таза сызбалық семантиканың болуына күмән туғызады. Сондықтан құрылымдық сызба негізінде сөйлем семантикасын модельдеу көптеген жағдайда зерттеушіні қанағаттандыра бермейді.

3.5 Антропоцентристік бағыт

Сөйлем мазмұнын зерттеу жолындағы жоғарыда аталған әдістерге тән қасиет – сөйлеуші мен тыңдаушы факторының қатыссыздығы, яғни коммуникант факторының ескерілмеуі. Бұл теория Ф. де Соссюрдің белгілі тезисіне сүйенеді,

яғни тілді іштей қарастыру және лингвистикадан сырт жағдайды ескермеу деген ұйғарым.

Қазір бұл тезис түгелдей өзгеріске ұшырап отыр. Коммуниканттар факторын ескермеу және коммуникация жағдаятымен санаспау сөйлем семантикасын анықтауда нәтижелі еместігі дәлелденді. Ал коммуниканттар мен коммуникация жағдайын ескеру "лингвистикалық прагматика" деген терминмен аталып жүр. Соңғы кезде лингвистикалық прагматика бағыты айтарлықтай дамып келеді.

3.5.1 Сөйлеуші факторы

Сөйлеушінің (жазушының), яғни коммуниканттың рөлі ең алдымен оның айтылым арқылы белгілі мақсатқа жетуге ниеттенуінде деп түсінген жөн. Сөйлеушінің коммуникативтік ниеті оның иллокутивті қызметі арқылы іске асады. Сөйлеушінің иллокутивті қызметі дегенді тыңдаушыға қалайша пропозициялық мазмұнды түсінудің нұсқауы деп те ұғынған жөн. Мұндай нұсқаусыз пропозициялық мазмұн ашылмай қалуы да мүмкін.

Қазіргі кезде әр автор әртүрлі иллоккутивті қызмет түрлерін ұсынуда. Солардың ішінде ең жиі кездесетіні:

1) ассертивтер – ұйғаруды, констациялауды, хабарлауды білдіреді;

2) директивтер – бұйрық ету, талап ету, өтініш ету, кеңес беру, ақыл айту;

3) декларативтер – үкім шығару, хабарландыру, бата беру;

4) комиссивтер – міндеттеме беру, сендіру, кепілдік беру, уәде беру, серт беру;

5) экспрессивтер – құттықтау, алғыс айту, тілек айту (білдіру), кешірім сұрау, өтінішін білдіру, т.б.

Әрбір иллокутивті қызмет сөйлеушінің белгілі коммуникативті ниетін білдіреді. Мысалы, комиссивтер бойынша, яғни міндеттеме беру, сендіру т.т. сөйлеушінің ниеті бойынша тыңдаушы (қабылдаушы) пропозицияда айтылған істің орындалуына кәміл сенуі қажет.

Иллокутивті қызмет басқаша да көрініс табуы мүмкін. Олардың ішінде тілекті, бұйрықты, сұрақты және т.б. білдіретін интонацияны немесе дауыс ырғағы арқылы берілуі де мүмкін. Сол сияқты рай категориялары арқылы және басқа да морфологиялық тәсілдер арқылы беріледі.

Иллокутивті қызмет сөйлеушінің ниетін білдірумен қатар ол тыңдаушымен (қабылдаушымен) де санасуды қажет етеді. Яғни тыңдаушының жекелік ерекшелігімен санаспай, сөйлеуші өзінің коммуникациялық мақсатына жете алмайды. Тыңдаушының қабылдау мүмкіндігін алдын ала білу, ол сөйлеуші адамның коммуникативтік мақсатқа жетуінің бірден-бір кепілі деуге болады.

3.5.2 Тыңдаушының моделі

Тыңдаушының моделі лингвистикалық әдебиеттерде әлі де толық жүйеге түспей, бірізді сипатқа ие болмай келеді. Дегенмен, коммуникация кезінде сөйлеуші өз мақсатына ойындағыдай жетуі үшін, ең алдымен, мынадай негізгі сипаттамаларды ескеруі қажет болады:

1) тыңдаушының (қабылдаушының) тілдік дәрежесі (компетенциясы), яғни тыңдаушының сөйлеуші тілін білу дәрежесі;

2) тыңдаушының әлеуметтік және мамандық сипаты;

3) тыңдаушының ұлттық сипаттамасы, дәлірек айтсақ, оның қатынас әлеміндегі мәдениеті мен этникалық нормасының деңгейі;

4) биологиялық жаратылысы: ер-әйел; үлкен адам, бала; денсаулығы жағынан – ауру-сырқау, сау; жарымжандық, кемтарлық және т.б. сипаттамалары;

5) психологиялық типі: циклотимдік-шизотимдік, интроверттік-экстраверттік; темпераменті: холерик, сангвиник, флегматик, меланхолик;

6) коммуникация кезіндегі психикалық қалпы;

7) тыңдаушының ұнату қабілеті мен әуестігі;

8) үй-жайы: үйленген-күйеуге шыққан, балалары бар не жоқ, ата-анасы және басқа туыстарының болу-болмауы;

9) коммуникация кезіндегі мақсаты мен қызығушылық сипаттары;

10) сөйлеуші мен тыңдаушының өзара таныстық дәрежесі.

Ешбір сөйлеуші тыңдаушының толық моделін біле бермейді, бірақ ол жайлы еш хабардар болмаса тілдік коммуникацияның әсері барынша әлсірейді.

Сөйлеушінің сөз арқылы тыңдаушыға ететін әсері оның ішкі дүниесінің, ішкі күйінің өзгеруіне әкеліп соғуы мүмкін. Мұндай өзгерісті **"перлокутивті эффект"** деген терминмен атауды ұйғарған.

Перлокутивті эффект сөйлеуші жағынан әдейі жоспарлы түрде немесе жоспарсыз, ойламаған жерден де болуы мүмкін. Қалай болғанда да, айтылым кезінде сөйлеу акты әртүрлі компоненттер арқылы іске асады. Олар мыналар:

1) айтылымның пропозициялық мазмұны (мыс., тыңдаушыға жағымсыз оқиға жайлы ақпарат жеткізу);

2) иллокутивті қызметтің түрі арқылы, яғни иллокутивті қызметті дұрыс таңдамау жағынан болатын әсер. Мысалы, мына сөйлемнің мағынасына көңіл аударалық: *"... маған бұйыруға сіздің қандай хақыңыз бар?"*. Шындығында, бұл жерде дауыс ырғағының дұрыс қойылмауынан, тыңдаушы сөйлеушіні дұрыс түсінбей қалу жағдайы немесе т.б. болуы мүмкін.

3) әр нәрсені орынсыз қабаттастыру арқылы. Мысалы, айтылымның жайлы дауыс ырғағы мен сазды түрде сипаттауы, синтаксистік құрылымның компоненттерді дұрыс не бұрыс таңдауында, жаңа және ескі ақпараттарды орны-орнымен дұрыс қоя білуінде және т.б.

Сөйлеушінің тыңдаушы моделін ескере білуіне қарай, өзара түсіністік немесе коммуникативтік қатынас туындайды. Соңғы кезде осындай жағдайды "кооперация" деп атауды ұйғарып жүр.

3.5.3 Тілдік қатынастағы кооперация принципі

Тілдік қатынас **"кооперация принципі"** деп аталатыны белгілі салт арқылы реттеліп отырады. Оның мәні мынада. Тілдік қатынасқа түсетін коммуниканттар, әдетте бірін-бірі түсінуге тырысады және сондай мақсатқа жету үшін өздерін де осыған лайықты ұстайды, яғни бір-бірімен серіктесуге ыңғай жасайды. Ал бұл **"серіктестік"** термині орысша **"кооперация"** деп аталып жүр.

Серіктестік (немесе кооперация) **принципі** ғалым Х.Грайстың (1985ж.) коммуникативтік постулаттары негізінде іске асады. Постулат – сипаттауды қажет етпейтін жорамал, дәлелсіз қабылданатын тұжырым. Коммуникативтік постулаттарға жататындар мыналар:

1) сандық (мөлшерлік) постулат немесе ақпарат мөлшерін анықтайтын – ақпараттық;

2) сапалық постулат немесе жалған ақпаратқа тыйым салатын – ақиқаттық;

3) айтылымның орындылығын білдіретін релеванттылық постулаты немесе тілдік бірліктерді айыруға қатысы бар мәнділік постулаты;

4) тәсіл постулаты немесе айтылымның бір мағыналығын қажет ететін айқындық постулаты.

Аталған постулаттар коммуникацияның идеалды тәсіліне жатады немесе, басқа сөзбен айтқанда, армандау (қиялдау) тәсілі. Сондықтан нақтылы тілдік қатынаста бұл аталғандар толық түрде көрініс таппайды. Бұған көптеген мысалдар келтіруге болар еді. Мәселен, тавтологиялық айтылымдарда сандық (мөлшерлік) постулат бұзылады – *әйелдің аты – әйел*; сапалыққа қатысты постулатында – қалжың түріндегі не астарлы айтылымдар. Мысалы, *"... себебі, ол өте ақылды ғой"* (*ақымақ қой* деген мағынада); айқындық постулатында – бұрып, күлдіргі айтылымдар (каламбур) – ауа райы, болжаушылардың айтуынша, *айдай ашық болып тұр* (шынында – *жауын құйып тұр*).

Сөйлеушінің айтылымдарындағы постулаттарда аталған кооперация (немесе серіктестік) принципінен ауытқулар әдейі қолданылуы да мүмкін. Осының нәтижесінде тыңдаушы тікелей мағынаға қатыссыз, постулаттарға қатысты ақпарат ала алады.

Осындай айтылымдағы анық түрде көрініс таппайтын ақпаратты **"импликатуралық"** ақпарат деп атайды. Бір айтылымның бірнеше **импликатурасы** болу мүмкін немесе ол айтылым **импликатуралық** жиынға ие болады деп пайымдайды.

Әрбір **импликатура** тілдік қатынастың нақты жағдайымен анықталады. Мысалы, жоғарыда келтірілген тавтологиялық *"әйелдің аты – әйел"* деген сөйлемдегі импликатуралардың мағыналары мынандай болуы мүмкін:

1) әйелдер – нашар, дәрменсіз, олақ және т.б.;

2) әйелдер – күшті, мықты, шыдамды, төзімді, берік, көнбісі және т.б.;

3) әйелдер – уақ-түйекшіл, ұсақ, қызғаншақ, күншіл, мансапқор, атаққұмар, менменсінген, даңғой және т.б.;

4) әйелдер – ісіне берілген, жан қияр, қамқор, қамқоршыл, ұқыпты, ықыласты, ілтипатты, зерделі, елегіш, сезгіш және т.б.

Мұндай мысалдарды іс жүзінде көптеп келтіруге болады.

Сонымен, сөйлемді прагматикалық контекст қатарына қосу оның мазмұнына белгілі анықтауды қажет ететін айтылымға айналдырады.

Сөйлем семантикасын модельдеуге антропоцентристік әдіс белгілі дәрежеде онтологоцентристік әдіске қарсы қойылады. Оның себебі антропоцентристік әдісте сөйлем мазмұны ішкі жағдаятқа, яғни субъективті факторға бағытталған, ал, екінші жағдайда – онтологоцентристікте сөйлем мазмұны сырт жағдайға, яғни объективті факторға, сыртқы әлемге бағытталған.

3.6 Құдайберген Жұбанов зерттеулеріндегі модельдеу әдісінің көрінісі

Профессор Қ.Жұбановтың* филологияға қосқан теориялық және практикалық ғылыми үлестері қазақ тіл білімінің барлық салаларына әсер етті деуге болады. Қазақ тілінің зерттелу болашағын өзіне тән көрегендікпен, табиғи дарындылықпен болжам жасап, олардың кейбір тұстарының ғалым зерттеулерінде орын алғанын арнайы атап кетуге болады. Ғалымның осындай зерттеу тұстары тілдің қолданбалы саласының жаңа бағыттарынан да байқалады. Дәлірек айтсақ, тіл зерттеу аясындағы XX ғ. 60-жылдарында ғана пайда болған "Статистикалық лингвистика", "Лингвистикадағы модельдеу" бағыттарының кейбір бастамаларын Құдайберген Жұбановтың сол кездегі зерттеулерінің өзінен-ақ молынан кездестіруге болады.

Қ.Жұбанов өткен ғасырдың 30-жылдарының өзінде-ақ өз зерттеулерін заманға сай жүргізуге ынталанғанын "Қазақ тілінің жоғарғы курсы" атты жазған еңбегіндегі мына тұжырымнан байқауға болады [Жұбанов Қ., 1999, 79-б.]: "XIX ғасырда жаратылыс ғылымы ілгеріледі; оларды зерттеуге түрлі метод қолданылды. Объектив әдісімен іс істеу тілшілерге де жұқты. Бұрын тіл логика негізді әдіспен зерттелетін еді (яғни, сағызша созыла беретін белгілі өлшеуге келмей, көбіне субъектив жорамалға сүйенетін)", – дейді.

Ғалымның сол кездің өзінде-ақ тіл білімінде сандық деректерді пайдаланудың қажеттілігіне де ерекше ден қойғаны байқалады. Мәселен, ғалымның 1936 жылы жарық көрген "Қазақ тілінің грамматикасы" атты еңбегінде сөз мүшелерінің орны, дыбыстардың саны мен сапасы, түрлері туралы жазған тұжырымы осыған дәлел: "Дыбыстың саны бар да, сапасы бар. Дыбыстың сапасы бірдей бола тұрып, саны әлденешеу бола береді. ... Бір тілде жұмсалатын дыбыстардың жалпы санының ұшы-қиыры жоқ, бірақ оның бәрінің бірдей сапасы әртүрлі бола бермейді. Сапасы әртүрлі болатын дыбыстар белгілі мөлшерде ғана болады да, көп болмайды. ... Қазақ тілінде де сапасы әртүрлі болатын дыбыстардың арнаулы саны бар..." [Жұбанов Қ., 1999, 183-б.]. Тілдік құбылыстарда сан мен сапаны ажырату қажеттігін айта келе, ғалым тұрпат (форма) пен мағынаны да ажырата білудің маңыздылығын тиянақтап айта білген. Мысалы, "Қазақ тілінің жоғарғы курсында" былай дейді [Жұбанов Қ., 1999, 79-б.]: "...Бірақ (грамматикалық құрылымдардың) сыртқы белгілеріне сүйенуден формализм туды. Сөйтіп морфология жеке сөздердің формасын, синтаксис сөз тіркестерінің формасын тексеретін ғылым деп танылды. ... Синтаксистің өзі де ... тіркескен сөздердің тіркесу формасын қарайтын ғылым болды" – дей келе, бұл әдісті толық қолдамайтынын білдіреді. Осындай күмәнін дәлелдей түсу мақсатымен және ол бағытты толығынан терістемейтінін білдіру

---

* Жұбанов Құдайберген Қуанұлы (19.12.1899 – 25.21938жж.) Ақтөбе обл., Темір ауд., Ақжар ауылында дүниеге келген. Қазақ тіл білімінің теориялық негізін қалаушы. Ленинградтағы (қаз. С-Петербург) Шығыстану ин-тының толық курсын 1928 ж. бітірген. 1930–1932 жж. КСРО ҒА Тіл білімі ин-тында аспирант. Түркітанушы, педагог-ағартушы. Қазақ филолог-ның тұңғыш профы. 1932–1937 жж. ҚазКСР Оқу халық комиссариатының ғыл. және оқу-әдістеме кеңесінің төрағасы, ҚазПИ-дің (қаз. Абай ат. Қаз.Ұлт.пед.унив.) қазақ тілі және әдебиеті каф. меңгерушісі, КСРО ҒА-ның Қазақстан бөлімшесінің сектор меңгерушісі, мемл. Терминкомның төрағасы. Республика халық ағарту комиссариатының коллегия мүшесі, КСРО ҒА Қазақстан бөлімшесі президиумының, бірнеше ғыл. кеңестер мен қоғамдардың мүшесі болған. Халық ағарту ісіндегі еңбегі үшін "Қазақстанның 15 жылдығы" белгісімен марапатталған. 1938 ж. "халық жауы" деген жаламен атылып, 1957 ж. КСРО Жоғ. Соты Соғыс коллегиясының үкімімен ақталды // Қазақ тілі энциклопедиясы. – Алматы, 1988. 146 бет.

үшін ғалым "Қазақ тілі грамматикасының кейбір мәселелері" атты еңбегінде тағы да былай дейді [Жұбанов Қ., 1999, 34-б.]: "...Рас, біз формашыларға қарсымыз: тіл құбылыстарын жалаң түр жағын ғана зерттеумен біліп болмайды. Бұл – сөзсіз сыңаржақтық. Сонда да жоғынан бары игі. Түр – форма деген тілде жоқ белгі емес, бар нәрсе. Бар қасиетін көрсетушілік қашанда пайда болмаса, зиян келтірмейді. Бірақ жеткіліксіз, өйткені түр (форма) тілдің бар тарауы емес, бір тарауы".

Шындығында да, ғалым зерттеулерінде тілдің формалық сипатына және объективтілікке сүйенетін жақтары баршылық.

Енді, қазіргі кезде қолданбалы тіл білімінің дәстүрлі бағытымен қатар дамып келе жатқан және заман сұранысына сай пайда болған оның жаңаша бағытындағы "тілді модельдеу" әдісінің жоғарыдағы тақырыпшаларда сөз болған кейбір теориялық тұстарын қайталай отыра, салыстырмалы пайымдаулар жасауға әрекет етейік.

Модельдеу әдісінің қажеттігі, ең алдымен, зерттеу нысанымызды тікелей байқауға, яғни қолмен ұстап, көзбен көруге болмайтын жағдайлардан туындайды. Сондықтан осындай модельдеу әдісі, ол өзінше бір бөлек әдіс ретінде, көптеген ғылымдар саласында, соның ішінде қазіргі кезде қазақ тілін зерттеуде де қолданыс тауып жүргені белгілі.

Ғылыми әдебиеттерде "модель" терминіне мынандай анықтама беріледі: "Модель дегеніміз – зерттеу нысанының қасиеттерін бойына сақтай алатын және түпнұсқаның (нысанның) орнына жүретін, танымдық процесінің қажеттігінен туындайтын материалдық не абстракты нысан (объект)" делінген.

Модель әр уақытта өзінің түпнұсқа түріндегі шындығынан да артық күйде, яғни "дәріптелген" сипатта қарастырылады. Осының негізінде, қай ғылым саласының болсын, формалды модельдері математикалық жүйе деп қарастырылады. Сондықтан, басқаша айтқанда, формалдылықтың түсінігі – математикалық дәлдік немесе бірізділік деуге болады. Бұл қасиеттер тіл табиғатына да тән құбылыс.

Осымен қатар, лингвистикалық модель эксплонаторлық қасиетке де ие болу керек деп ұйғарылады, яғни оның түсініктеме беру мүмкіндігі болуы қажет. Аталған қасиетке ие болуы үшін, ол, ең алдымен, осыған дейінгі тілдік теория негізінде түсінігі күңгірт не түсініксіз болып жүрген тілдік құбылыстарды бірізді түсіндіре алатын, анықтама бере алатын мүмкіндігі болуы керек. Сол сияқты, зерттеу нысанының бұрын-соңды ескерілмей жүрген жаңа мүмкіндіктеріне де ғылыми болжам жасай алатын қабілеті болуы қажет деп пайымдалады.

"Лингвистикалық модельдеуде" сөйлеу қызметі моделінің пішінін екі түрде қарастырады. Оның біріншісі – семантикасыз немесе тек грамматиканы игеруге ғана "еліктеуші" таза синтаксистік модель. Екіншісі – мағыналы сөйлемдерді түсіндіре білумен қатар, оларды дұрыс құра да білу мүмкіндігіне еліктейтін – семантикалық модельдер.

Мұндай модельдерді құру жазба не сөйлеу мәтіндерін автоматты түрде өңдеуде аса маңызды. Мәселен, "машиналық аударма" жүйелерінің барлық дамыған түрлерінде "автоматтандырылған синтаксистік талдау" жүргізу сөйлемнің семантикалық түсінігін алу үшін және екінші тілдік жүйедегі сөйлемдер құрылымын "аударма грамматикасына" ауыстыру үшін аса қажет деп саналады. Сол сияқты, "ақпараттық іздеу" жүйесінде де толық синтаксистік талдау сөйлем құрамындағы "есімді сөз тіркестері" негізінде танылады. Ал

"жасанды интеллект" жасау мәселесіндегі мәтінді тану жүйесіне қатысты синтаксистік талдау, өзінше бір бөлек семантика-синтаксистік топтаманы құрайды деуге болады.

Қазіргі кезде, қай тілдің болмасын, оған тән сөйлем құрылымының дәл және формалды берілісін білу аса қажет болуда. Ол үшін реттелген сызықтық тізбек құрайтын кесінділер жиыны – **"құрастырушылар жүйесін"** немесе **"құрастырушы-лар моделін"** құра білу керек. Аталған модельде синтаксистік бірлік ретінде сөзтұлғалар, сөз тіркестер және олардың құрылымдық қатынастары қарастырылады.

Жоғарыдағы тақырыпшаларда баяндалғандай, жүйе жиынындағы қатынастар "граф" деп аталатын сызба түрінде көрініс табады да және оның төбешігіндегі ең кіші бірлік ретінде не негіз сөз, не сөзтұлға, не сөз тіркес алынады. Сөйлемдердің графтық сызбасын салудың алдында сөйлемдегі әрбір сөзқолданыстар өз реті бойынша нөмірленеді. Әрі қарай, сызбалар түрінде граф төбешігіндегі құрастырушы сөзқолданыстар мағыналық жағынан ішкі сөз тіркестерге бөлектеліп, олар бір-бірімен ішкі және сыртқы тәуелділіктеріне қарай, түзу сызықтық кесінділер арқылы өзара қосылады. Осындай сызба пішіні әдебиеттерде **"құрастырушылар моделі"** немесе **"синтаксистік модель"** деген терминдермен аталып жүр.

Сөйлем құрастырушыларына талдау жүргізудің өзіндік қиыншылығына қарамай, **"құрастырушылар модельдері"** сөйлем құрылымын грамматикалық және семантикалық тұрғыда да дұрыс сипаттайды деп саналады. Мұндай дұрыстық сипат, көбінде, мағына ажырату кезінде, мысалы, құрастырушы сөзтұлғалар – **"омонимдес сөздер"** немесе сөйлемдер – **"омонимдес сөйлемдер"** болған жағдайларда анығырақ байқалатыны дәлелденеді.

Сөйлемдер ішіндегі сөзтұлғалар арасындағы **қабысу, жанасу, қиысу, меңгеру** сияқты тәуелділіктер жиыны сол сөйлемнің синтаксистік құрылымы жайлы көптеген мәліметтер бере алады. Ал тәжірибе жүзіндегі ізденістерде, мұндай тәуелділіктердің жиыны граф пішініндегі сызба түрінде көрініс табады.

Профессор Құдайберген Жұбановтың сөйлем құрылымына қатысты зерттеулерінде сөйлем ішіндегі сөзтұлғалар мен сөз тіркестер "граф" немесе "құрастырушылар моделі" деген терминдермен аталмаса да, сөйлемге енетін сөздердің әрбіреуін тиісті нөмірлермен таңбалау және олардың тәуелділіктерін сызба түрінде кесінділер мен төртбұрыштар арқылы беру молынан кездеседі.

Мәселен, ғалымның "О формах сочетания слов в казахском языке" атты мақаласында [Жұбанов Қ., 1999, 476-б.] 15 сөзқолданыстан тұратын мындай сөйлемді қарастырады: *"Сен(1) көрген(2) қымбат(3) алпауыттан(4) да(5) Қызыл(6) құрт(7) Ертай(8) ұстаның(9) Темірбайдан(10) алған(11) төрт(12) бесті(13) аты(14) жақсы(15)"*. Осы сөйлемдегі сөздердің орын ауысуынан сөйлем мағынасының өзгеру заңдылығын көрсету мақсатымен, әрбір сөзді нөмірлеп алып (1-суреттегі бір сөйлем), бес түрлі айтылым ретінде бөлек-бөлек қарастырып, пайымдаулар жасайды. Сөйлемді орыс тіліне былайша аударады: *"Чем(4) дорогой(3) помещик(4) виденный(2) тобою(1), лучше(15) и(5) гнедой(12) пятилетний(13) мерин(14), полученный(11) кузнецом(9) Ертаем(8) кызыл-куртовцем(6) от(7) Темирбая(10)"*.

Әрі қарай осы сөйлемдегі сөздерді түпнұсқадағы баламаларымен салыстыра отыра, олардың мағыналары мен тұлға өзгерістерін талдайды. Міне, мұндай әрекеттер

қазіргі кезде арнайы әдіс ретінде сөйлем синтаксисін модельдеуге қатысты зерттеулерде жиі қолданыс тауып жүр.

Қ.Жұбановтың зерттеулерінде [Жұбанов, 1999], яғни **"Сөйлем мүшелері"** [153-157-бб.], **"Жетек сөз бен жетекші сөз"** [160-161-бб.], **"Сөйлем мүшелерінің орын-орны"** [162-б.] деген тақырыпшаларда **"О формах сочетания слов в казахском языке"** [476-481-бб.], **"О построении речи на казахском языке"** [482-495-бб.] және **"Этюды к переводу трех стилей"** [531-534-бб.] атты мақалаларында көрініс тапқан сызбалар үлгісі 12 түрлі сурет арқылы берілген. Бұл келтірілген сызбаларды қолданбалы лингвистика терминімен айтсақ, белгілі мағынада, **"Қазақ сөйлемдерінің синтаксистік модельдерінің үлгілері"** деуге де болады *(1-12 суреттерді қара).*

Енді сөйлем ішіндегі тәуелділіктердің кесінді сызба түрінде көрініс табуын Қ.Жұбановтың аталған кітабындағы мынадай пайымдауларынан байқауға болады:

1) кітаптың 156-бетінде *"Айқындауышты өзінің айқындап тұрған сөзімен сызық арқылы байланыстырып көрсетсек мынадай болады"* деп (төменде көрсеткен 2-ші және 3-ші суреттегідей) екі түрлі сызбаны ұсынады;

2) кітаптың 160-бетінде – *"Біздер мектепке келдік"* деген сөйлемді мысал ретінде алып, былай пайымдаған: "... бұл үш сөз матастырылмаған. *Біздер* деген мен *мектепке* деген екеуі жиналып *келдік деген бір сөзге арқандалған. Мұның арқандалу ретін сызықпен көрсетсек мынадай болады" деп төменде көрініс тапқан* 4-суретті *және "Енді қай сөз, қай сөзді жетектеп тұрғанын да сызбамен таңбаласақ, мынадай болады"* деп келесі *5-суретті* ұсынады;

3) кітаптың 162-бетіндегі "Сөйлем мүшелерінің орын-орны" деген тақырыпшада бастауыш, баяндауыш, анықтауыш және толықтауыш болатын сөйлем мүшелерінің орындарын көрсету үшін М.Әуезов шығармасынан алынған мына сөйлемнің *"Бұл – мал күткен, малға ықылас берген бейілдің көрмесі"* әр сөзін нөмірлеп алып, былай дейді: "Мұндағы сөздердің қалайша қиындасып тұрғанын сызықпен көрсетсек, мынандай болады" деп төмендегі *6-суретте* көрініс тапқан сызбаға сілтеме жасайды.

4) кітаптың 482 бетіндегі "О построении речи на казахском языке" атты мақаласында 16 сөзқолданыстан тұратын сөйлемді қарастырып, олардың формалды белгілері арқылы және бір сөзге қатысты параллель тіркестердің орын тәртібінің қатал сақталмай, тек логикалық екпінге ғана бағыныштылығы жайлы сызбаны ұсынады. *7-суретте* орын алған бұл сызбаның сөйлемнің синтаксистік моделінің көрінісі екендігіне осы саланың мамандары күмән туғызбас деген ойдамыз.

5) келесі қарастыратынымыз ғалым кітабының 531-бетіндегі көлемі жағынан шағын (4 бет), ал құндылығы жағынан "аударманың синтаксистік және семантикалық моделі" ғылыми бағыты үшін аса маңызды "Этюды к переводу трех стилей" атты (тақырыбы шартты – *ред.)* атты зерттеу жұмысы. Бұл мақала ҚазССР-ның Орталық архивінен алынып, тұңғыш рет Алматыда "Лексико-семантические вопросы перевода" атты жинақта 1987 жылы басылып шыққан болатын. Мұның құндылығы үш түрлі стиль мәтіндеріндегі (көркем-әдебиет, ғылыми және публицистикалық) сөйлемдерді салыстыра-салғастыра қарастыруында. Әр стильге тән ерекшеліктерге қатысты сөйлемдер ішіндегі сөздердің бір-бірімен байланысын 5 түрлі сызбалар арқылы орыс тіліндегі сөйлемдерді қазақ тіліне аударуға қатысты өзіндік семантика-синтаксистік модель-үлгіні ұсынуында. Аталған сызбалар *8-12-суреттерде* көрініс тапты.

Ғалым Құдайберген Жұбанов аталған сызбаларына сілтеме жасау үшін мынандай айтылымдарды қолданған: *"сызық арқылы байланыстырып көрсетсек", "мұның арқандалу ретін сызықпен көрсетсек", "қай сөз, қай сөзді жетектеп тұрғанын да сызбамен таңбаласақ", "сөздердің қалайша қиындасып тұрғанын сызықпен көрсетсек"* және т.б.

Тілді оның формалды пішіні арқылы сипаттау "алмасу ақпаратының" мазмұндық сырын ашуға толық мүмкіндік туғыза бермейтіндіктен зерттеуге тиісті ақпараттың мағыналық, яғни семантикалық жағын бөлек қарастыру қажеттігі туындайды.

Жоғарыдағы тақырыпшаларда сөз болғандай, көптеген зерттеушілердің тұжырымдауынша, сөйлем мағынасы аса күрделі және көпаспектілі құрылым. Сондықтан синтаксистік модельдеу барысында ұстаным етілетін төрт түрлі ғылыми бағыттың үшеуі "әлем—ойлау—тіл" үшбұрышы элементтерінің қатыстылығына қарай, ал төртіншісі – "тіл—адам" прагматикалық қатынасына негізделеді. Ғылыми әдебиеттерде бұл бағыттар: *онтологоцентристік, концептоцентристік, синтактикоцентристік* және *антропоцентристік бағыт* деген терминдермен аталып жүр.

Қазіргі кезде *онтологоцентристік* және *концептоцентристік* бағыттар бойынша коммуниканттар факторын ескермеу және коммуникация жағдайымен санаспау сөйлем семантикасын дұрыс ашуда нәтижелі болмайтындығы да дәлелденді. Сондықтан бұл жағдайларды ескеру "лингвистикалық прагматика" деген терминмен аталады. Айтылым мазмұнына семантикалық компонентпен бірге прагматикалық компоненттің де енуін қажет деп санайды. Оған жататындар: сөйлеуші факторы, тыңдаушы факторы және сөйлеу актісіне қатысы бар – орын мен уақыт, сөйлеуші мен тыңдаушының әлемдік көзқарасы, жалпы білім қоры және т.б.

Ғылыми әдебиеттерде сөйлеушіге қатысты айтылым үлгілері – "сөйлеуші моделі", ал қабылдаушыға не тыңдаушыға қатысы тілдік мәселелер – "тыңдаушы моделі" деген терминдермен (модель говорящего, модель слушающего) аталып жүр. Сөйлеуші-коммуниканттың рөлі, ең алдымен, айтылым арқылы белгілі мақсатқа жетуге ниеттену үшін иллокутивті функцияларды пайдалануды қажет ететіні белгілі. Ал иллокутивті функция сөйлеушінің ниетін білдірумен қатар ол тыңдаушымен де санасуы қажет етеді екен.

Міне, осындай *(онтологоцентристік, концептоцентристік, синтактикоцентристік, антропоцентристік* және т.б.) толып жатқан терминдер арқылы аталып жүрген "қолданбалы тіл білімінің" жаңа бағытына қатысты мәселелер Қ.Жұбановтың зерттеулерінде дәл олай аталмағанымен, өзіндік мазмұнда молынан көрініс тапқан.

Жоғарыда қысқаша баяндалған мәселелер қазақ тілінің қолданбалы саласының жаңа бағыты бойынша жүргізілетін зерттеулерде өз жалғасын табатыны мәлім.

Сонымен, қазақ тілінің теориялық және қолданбалы тіл білімінің негізін қалаушы, қазақ филологиясының тұңғыш профессоры Құдайберген Қуанұлы Жұбановтың азғана өмірінің бір-ақ мақсаты болған сияқты. Ол – әлі ашылмаған шындықты ашу немесе соның ашылуына ықпал жасау, сол арқылы қоғамға қызмет ету. Мұндай мақсатқа жетудің әртүрлі жолдары бар екені белгілі, бірақ ғалым Қ.Жұбанов солардың ішінен зерттеудің дәлдігі мен объективтілігін қамтамасыз ететін тәсілді ғана ең дұрысы деп тапқан болу керек. Шындығында да, дәлдік пен объективтілікке ұмтылу – қазіргі ғылымдарға бірден-бір тән қасиет, ал бұның тіл біліміне де қатысы бары сөзсіз.

1 2 3 4 5 6 7 8 9
I. Сен көрген қымбат алпауыттан да Қызыл құрт Ертай ұстаның

10 11 12 13 14 15
Темірбайдан алған төрт бесті аты жақсы.

1 2 3 4 5 6 7 8
II. Қызыл құрт Ертай ұстаның Темірбайдан көрген аты қымбат.

9 10 11 12 13 14 15
Торы алпауыттан да сен алған бесті жақсы.

1 2 3 4 5 6 7
III. Ұстаның жақсы көрген аты – алпауыт қызыл бесті.

8 9 10 11 12 13 14 15
Сен Темірбайдан алған құрт Ертай торыдан да қымбат.

1 2 3 4 5 6 7 8 9
IV. Қызыл құрт Ертай ұстаның алпауыттан да қымбат көрген аты

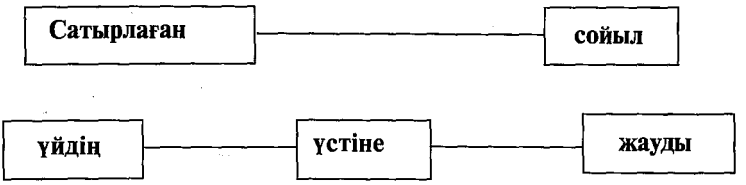
10 11 12 13 14 15
Темірбайдан сен алған жақсы торы бесті.

1 2 3 4 5 6 7 8 9
V. Темірбайдан алған бесті жақсы сен көрген Қызыл құрт Ертай.

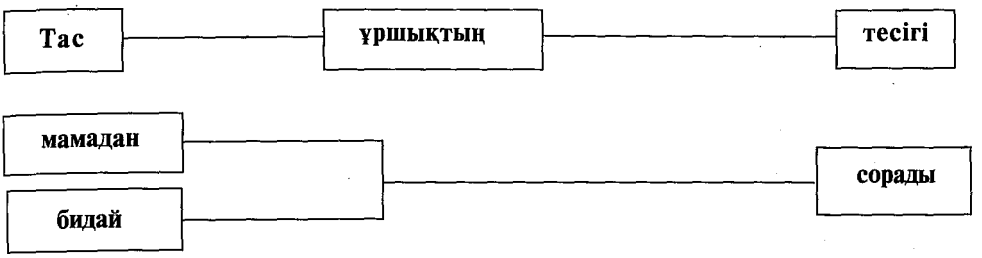
10 11 12 13 14 15
Торы алпауыттан да ұстаның аты қымбат.

1-сурет

"Сатырлаған сойыл үйдің үстіне жауды"



"Тас ұршықтың тесігі, Мамадан бидай сорады". **(Ілияс)**



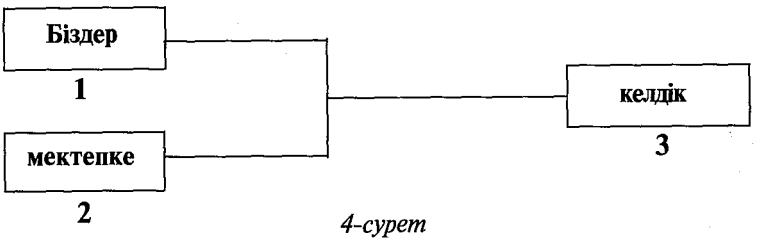
3-сурет

Сөйлем ішінде сөздер бір-бірін ерте де, біріне-бірі ере де байланысады. Еруші сөзді **жетек сөз**, ертуші сөзді **жетекші сөз** дейміз.

1 2 3

Мысалы: ***Біздер мектепке келдік.***

Мұның арқандалу ретін сызықпен көрсетсек, мынандай болады:



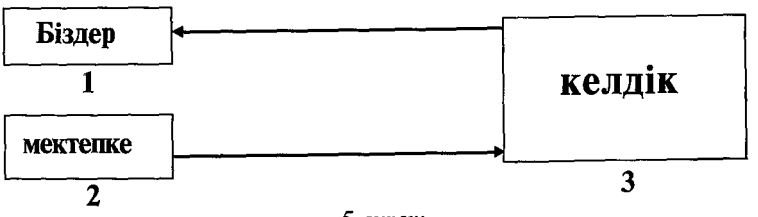
4-сурет

Сөйлем ішінде сөздер бір-бірін ерте де, біріне-бірі ере де байланысады. Еруші сөзді **жетек сөз**, ертуші сөзді **жетекші сөз** дейміз.

1 2 3

Мысалы: ***Біздер мектепке келдік.***

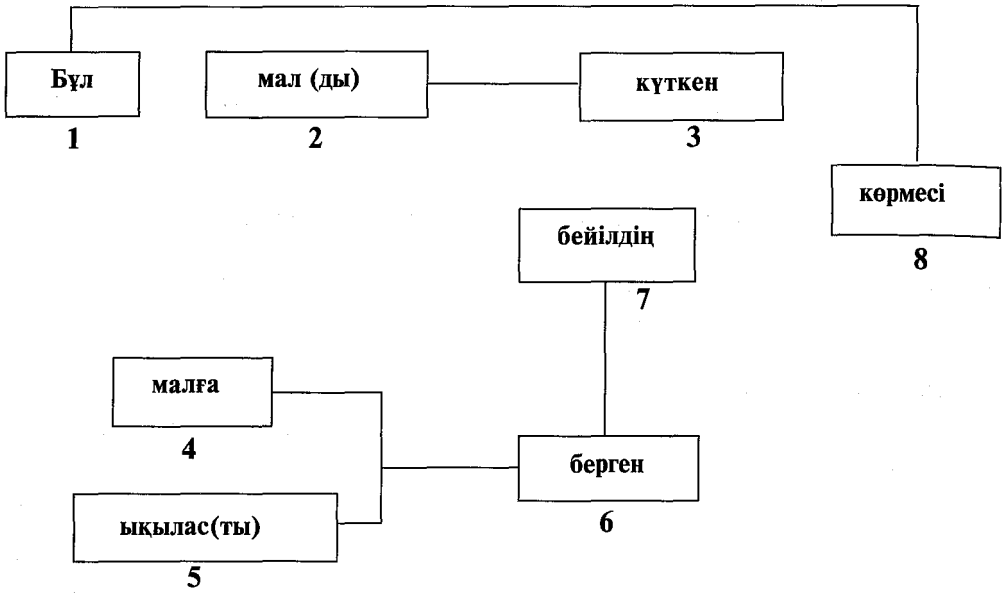
Енді қай сөз, қай сөзді жетектеп тұрғанын да сызықпен таңбаласақ, мынандай болады:



5-сурет

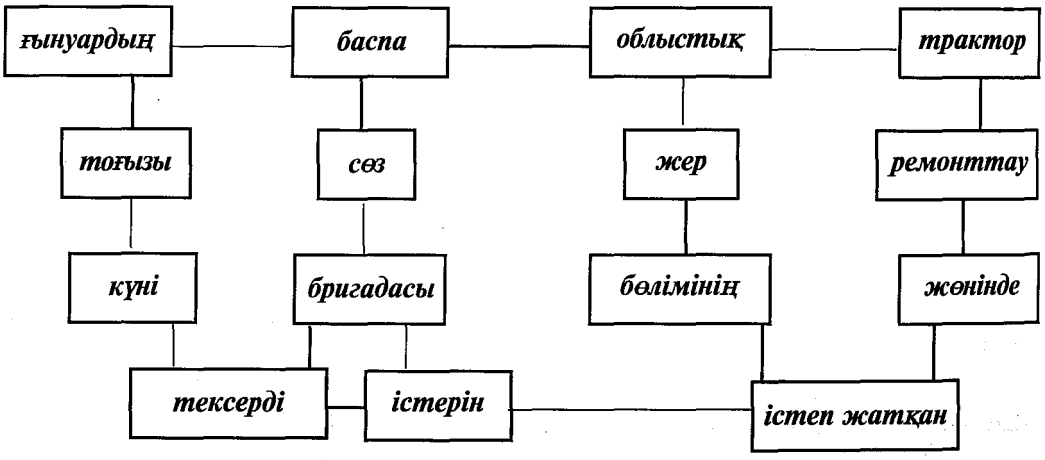
1 2 3 4 5 6 7 8

Бұл – мал күткен, малға ықылас берген бейілдің көрмесі. **(Мұхтар)**



6-сурет

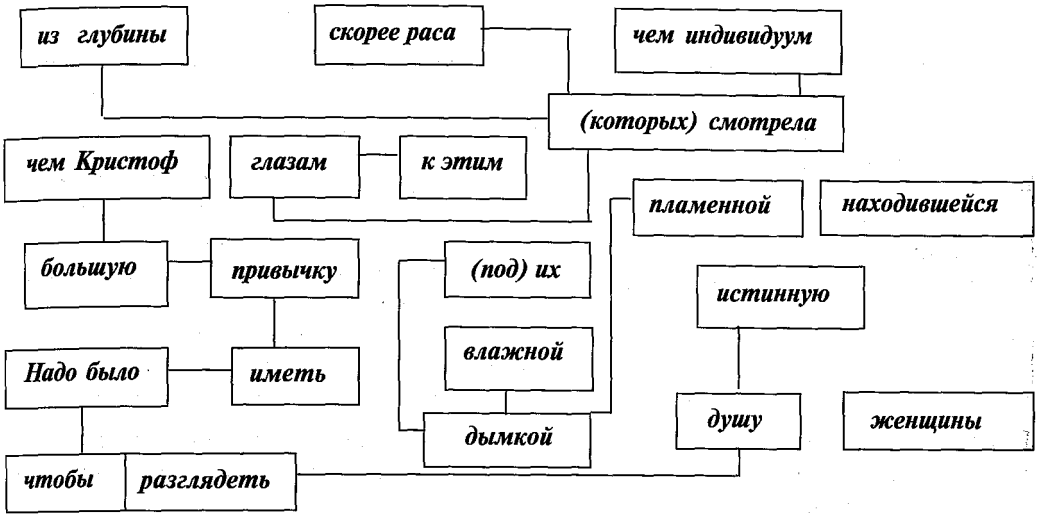
"Трактор ремонты жөнінде облыстық жер бөлімінің істеп жатқан істерін баспасөз бригадасы ғынуардың тоғызы күні тексерді".



7-сурет

I. Текст художественный (Ромен Роллан. Жан Кристоф. Т.4. С. 68). ***а) "Надо было иметь большую, чем Кристоф, привычку к этим глазам, из глубины которых смотрела скорее раса, чем индивидуум, чтобы разглядеть под их влажной и пламенной дымкой истинную душу находившейся перед ним женщины".***

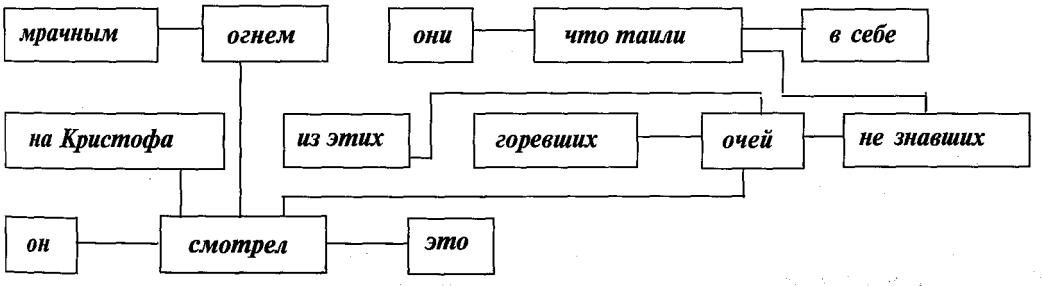
Схема связей слов



8-сурет

б) ***"Это он смотрел на Кристофа из этих горевших мрачным огнем очей, не знавших и сами, что они в себе таили".***

Схема связей слов



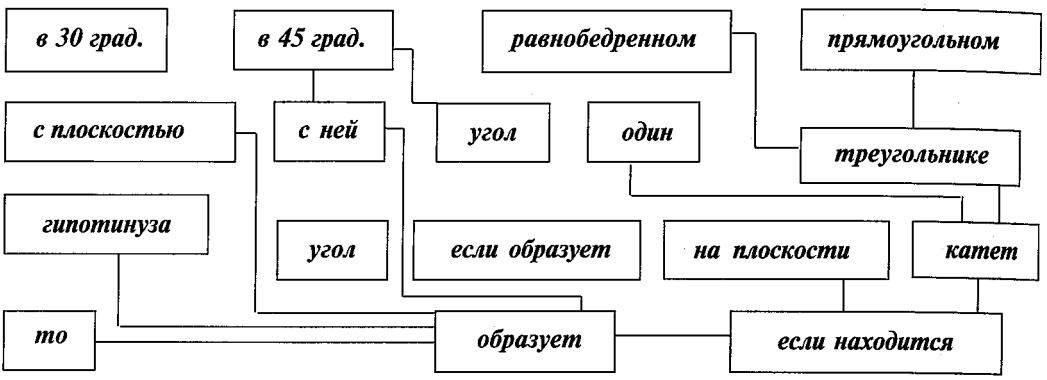
Перевод: ***"Өзінің ішінде не бар екенін өзі білмей, оттай жанып, түнерген ол екі көзбен Кристофқа қарап отырған сол бүтін халық еді".***

9-сурет

II. Текст научный (Рыбкин Н. Сб. задач по геометрии. Ч.2. Стереометрия. М., 1934. С. 7, задача 12).

"Если в равнобедренном прямоугольном треугольнике один катет находится на плоскости М, а другой катет образует с ней угол в 45 градусов, то гипотенуза образует с плоскостью М угол в 30 градусов".

Схема связей слов



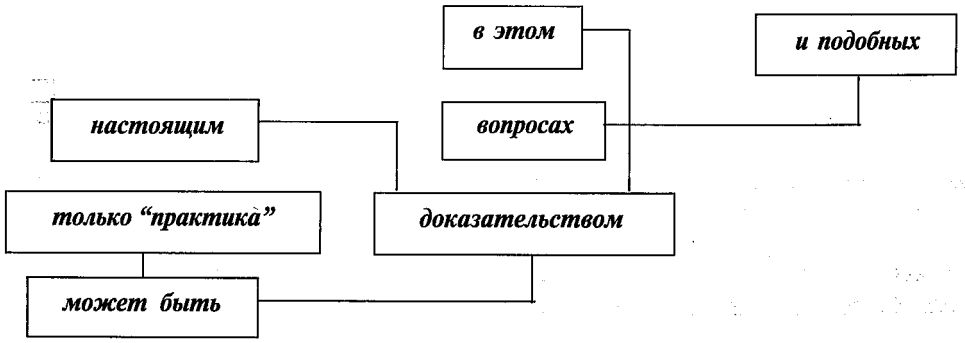
Перевод: ***"Тең бүйірлі тік бұрышты үшкілдің бір катеті М жазықтығында да, екіншісінің ол жазықтықпен арасы 45 градус болса, онда гипотенуздың жазықтықпен арасы 30 градус болады".***

10-сурет

III. Текст публицистический (из газет).

а) "Настоящим доказательством в этом и подобных вопросах может быть только "практика".

Схема связей слов

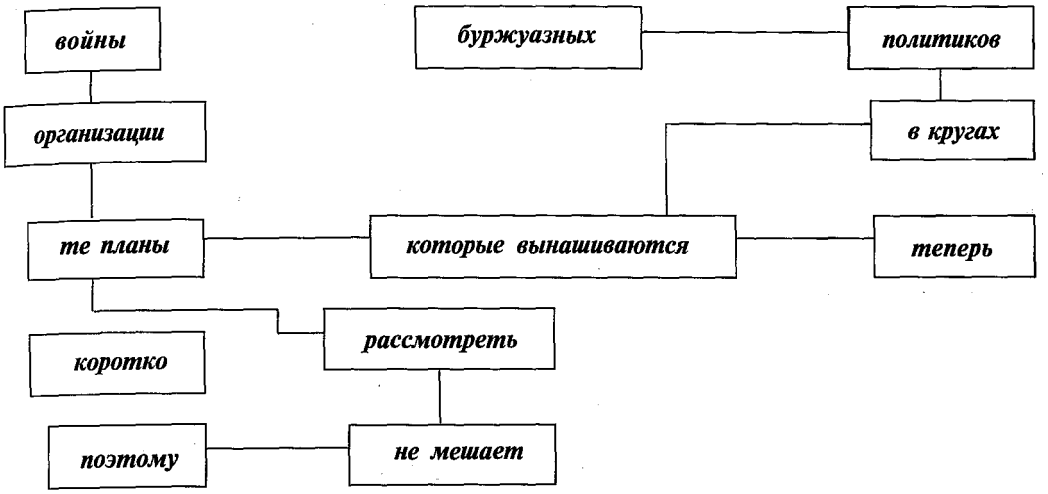


Перевод: ***"Бұл мәселеде, бұл сияқты мәселелерде нағыз "ыспат" тек тәжірибеде қана бола алады".***

11-сурет

б) "Не мешает поэтому рассмотреть коротко те планы организации войны, которые вынашиваются теперь в кругах буржуазных политиков".

Схема связей слов



12-сурет

Төртінші тарау
ҚОЛДАНБАЛЫ ЛИНГВИСТИКАНЫҢ ФОНЕТИКАЛЫҚ АСПЕКТІЛЕРІ

4.1 Тілдік бірліктердің дыбыстық тұрпатын зерттеудің негізгі бағыттары

Кез келген табиғи тілдің негізі мен бастамасында тіл қызметінің дыбыстық тұрпаты жататыны белгілі. Міне, осы жайттың қолданбалы тіл біліміндегі фонетикалық зерттеулер жүргізуде рөлі айрықша болып саналады.

Қазіргі кезде тілдік бірліктердің дыбыстық тұрпатын зерттейтін негізгі бағыттарына мыналарды жатқызуға болады [Бондарко, 1996, 113-141-бб.]:

1. Сөйлеу сигналының акустикалық сипаттарын талдау.

Бұл зерттеу бағытының тәжірибелік рөлі мынада:

- байланыс арнасы бойынша таралатын тілдік ақпарат жүйесін құру;
- сөйлеу тілін автоматты түрде тани білетін және жинақтай (синтездей) алатын мәселелерді шешу;
- сөйлеу мен есту қабілеттеріне зақым келген ауру адамдардың мүмкіндіктерін арттыруға бағытталған әдістемелерді жетілдіру.

2. Сөйлеуді туындатуға қатысты процестердің артикуляторлық сипаттарын талдау.

Мұндай зерттеу арқылы тиісті ми қабы орталықтары жағынан дыбыстау мүшелеріне қатысты басқару тетігін анықтау. Бұл адам баласының жоғарғы нерв қызметін сипаттайтын процестерді ұғынуда аса маңызды. Осыған орай, бұл зерттеу бағытының сөйлеудің әртүрлі формалары бұзылған ауру адамдарды жазуда да практикалық маңызы зор. Көптеген зерттеушілердің тұжырымдауынша, сөйлеуді туындату тетігі жайлы мәліметтер сөйлеу тілін тану мен жинақтаудың (синтездеудің) автоматтанған жүйесін құру үшін маңызды.

3. Сөйлеу түрінде берілген хабарламаның адамның есту жүйесінің қабылдау мүмкіндігін зерттеу.

Мұндай зерттеу, біріншіден, хабарламаның акустикалық қасиеттерінің құндылығын бағалауда, екіншіден, сөйлеу сигналын тануда, өңдеу моделін құруда аса маңызды. Сонымен бірге зерттеу мәліметтері сөйлеу түрінде берілген хабарламаны жазу мен тарату аясында қызмет істейтін мамандарға және тілді автоматты түрде тану мен жинақтаумен айналысатын мамандарға айтарлықтай маңызды деп саналады.

Сөйлеу түріндегі сигналдардың сыр-сипатын зерттеу ісінің ана тілі мен шет тілдерін оқытуда да маңызды екендігіне айырықша көңіл бөлу керектігін баса айтпақпыз. Бұл аталған мәселеде орфоэпиялық дағдыны, диалектілік және қарапайым сөз қолдану құбылыстарын жою, сондай-ақ кейбір сахналық айтылымдарды қалыптандыру және жолға қою қажеттігі туындайды. Енді осы айтылғандарға байланысты қолданбалы фонетика саласының негізгі типтік жағдайларын қарастырайық.

4.2 Байланыс пен хабар тарату арналарындағы сөйлеу тілінің сигналы

Акустикалық сигналдарды жазу мен тарату аясындағы мамандар акустикалық сигналдарды *бірінші реттегі* және *екінші реттегі* деп екіге бөледі. Біріншілері – ол сөйлеу сигналдарының өзі, ал екіншілері – электроакустикалық құрылғылар арқылы шығарылатын сигналдар. Яғни бірінші реттегі сигналдар электроакустикалық байланыс және хабарлау желісінен өтіп, оған сәйкес параметрлері өзгеріске ұшырап, екінші реттегі акустикалық сигналға ауысады деп ұғыну керек [Сапоржников, 1978, 40-б.]. Акустикалық өлшемдердің өзгерісі сигналдың динамикалық пен жиіліктік кеңістігіне, спектрлік және уақыттық сипаттамаларына қатысты болады. Хабарды жазу мен тарату жүйелерін құрушылардың негізгі міндеті тыңдаушыға екінші реттегі сигналдың дәл бірінші реттегі сигналдай әсер қалдыруына талпыну. Әрине, бұл айтылғандар әр уақытта мүмкін бола бермейді. Сондықтан екінші реттегі сигналдың сапасын жақсарту жолында әртүрлі объективті және субъективті әдістер қолданылып жүр.

Аппарат арқылы жасалған екінші реттегі сигналдың жиіліктік, динамикалық, уақыттық сипаттарын адамның сөйлеу кезіндегі осындай сипаттамаларымен объективті түрде салыстыру. Осы арқылы екінші реттегі сигналдың сапасын арттыру. Мысалы, егер біз адамның сөйлеу кезіндегі динамикалық диапазонының аралығы *25-35 дБ* екенін біле тұра, хабарламаны жазып алатын микрофонның динамикалық диапазонын *20 дБ* мөлшерінде алсақ, біз онда екінші реттегі сигналдың бірінші реттегі табиғи сөйлеу түріндегі сигналдан барынша осал болатынын алдын ала айта аламыз.

Жеке адамның есту мүмкіндігі арқылы алынған екінші реттегі сигнал мәліметтерін бағалау – *субъективті әдіс* деп аталады. Әрине, алынған мәліметтерге сүйенетін болсақ, тәжірибеге қатысқан жеке адамның есту қабілетіне немесе тәжірибе шартына тәуелді болмаған жөн. Сол үшін сынауға тиісті дыбыстық нысан мен қатысушы адамдардың саны барынша көп болуы керек. Міне, осындай жағдайда ғана объективті нәтижеге жетуге болады. Аталған тәжірибелер негізінде байланыс желісі арқылы таратылатын сөйлеу тілінің айқындығы анықталады. Дұрыс қабылданған сигналдар санының барлық таратылған сигналдар санына қатынасын білдіретін шаманы *айқындық өлшемі* деп атайды.

Тілімізде қолданылып жүрген сигналдар туралы сөз болғанда, оны дұрыс тану нәтижесі көбіне таралатын сигналдар түріне байланысты. Айталық, байланыс желісі өте "жаман" (төмен) сапада болсын делік. Яғни оның өткізу жолы өте тар және үлкен кедергі келтіреді деген сөз. Егер осындай байланыс желісі арқылы сөздер немесе сөйлемдер таратылатын болса, оның дұрыс танылуы тек қана тәжірибеге қатысқан адамның жоғары дәрежелі есту мен болжау қабілетіне байланысты болады. Адам баласының дыбысты қабылдау қабілетінің өте жоғары дәрежелі болуынан, олар жартылай немесе бұрмаланған акустикалық сигналдарды дәл танып және оған тиісті баға беруі қиын. Сондықтан екінші дәрежедегі сигналдардың дұрыс-бұрыстығына сын айту үшін, ең алдымен мағынасыз дыбыс тіркесімдеріне жататын буындарды немесе жалған сөздерді танып-сынау тәжірибесін жүргізу қажет болады.

Нақты бір тілге қатысты дыбыс тізбелерінің статистикалық сипаттамаларын анықтау – ең бір маңызды фонетикалық мәселе. Себебі жалған және дұрыс буындар мен сөздердің анық түрде дыбысталуларының статистикалық заңдылықтары бірдей. Осындай қолданбалы есептердің әртүрлі тілдер материалдары бойынша шешімін

іздестіру арқылы ғалымдар көптеген статистикалық зерттеулердің түрлерін өмірге келтірді. Аталған зерттеулер негізінде, сөйлеудің айқындығын өлшеу нәтижелерінің қоры жинақталды. Сонымен бірге, аталған статистикалық мәліметтердің маңыздылығы тілді типологиялық тұрғыда қарастыру кезінде де арта түсті.

4.3 Тіл дыбыстарының негізгі акустикалық сипаттамалары

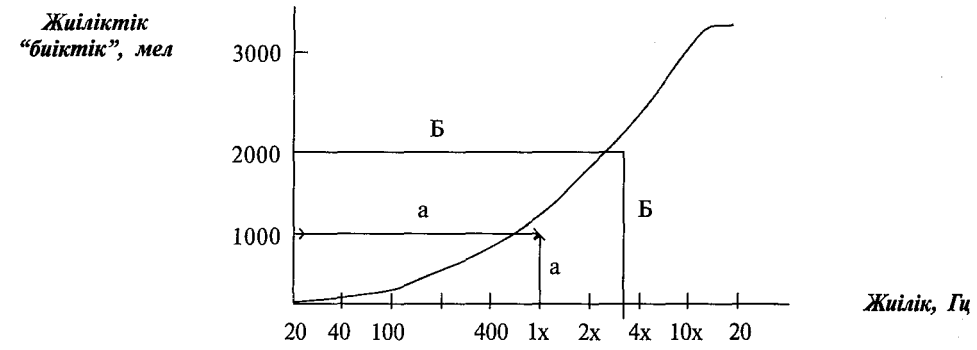
Табиғат дыбыстары сияқты сөйлеу кезіндегі естілетін дыбыстар да ауа ортасының тербелгіш қозғалыстарының нәтижесінде адамның есту аппаратына тигізетін әсерінен туындайды. Мұндай тербелгіш қозғалыстардың негізгі белгілері төмендегідей сипатталады.

1. Тербелгіш қозғалыстардың жиілігі уақыт бірлігі ішіндегі тербелгіш қозғалыс санымен анықталады. Жиіліктің бірлік өлшемі неміс физигі Герцтің құрметіне "герц" (Гц) терминімен аталады.

Егер қайсыбір дыбыстың жиілігі *100 Гц*-ке тең болса, оны *1* секунд ішіндегі дыбыстың үзіліссіз тербеліс қозғалысы *100* ретке дейін жетеді деп түсіну керек. Адамдар дыбыстың жиілік мөлшеріне қарай, жиіліктің жоғары не төмен болу сипатын (жоғарылық сипатын) сезінеді. Мысалы, дыбыстың жиілігі ұлғайған сайын, адамның есту жүйесіне оның тигізетін әсері жоғары сипатта болады.

Дыбыстың жоғарылық бірлігі "мел" деп аталады. Арнайы әдебиеттерде жиілікті латынның *f* әрпімен белгілейді (ағылшынша - *frquency*).

Дыбыстың жиілігі мен жоғарылығы арасындағы қатынас *1-суретте* көрсетілді.

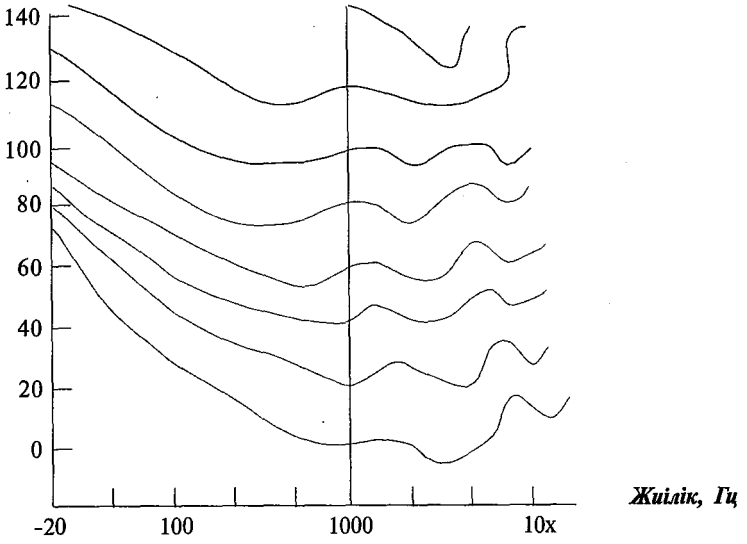


1-сурет

2. Дыбыстың қарқындылығы (интенсивтілігі) немесе оның күші тербелгіш қозғалыстың амплитудасымен өлшенеді. Ауамен таралатын дыбыстардың тербелісін ауаның қоюлануы мен сиретілуінің кезектесіп келетін тізбегі деуге болады. Яғни, басқаша айтқанда, ол – атмосфералық қысыммен салыстырғандағы ауа қысымының өзгерісі. Мұндай қысым аудан бірлігіне әсер етуші күш шамасына тең. Қарқындылыққа тәуелді болатын дыбыстың күшті естілуі дыбыс қысымының артуына тікелей байлыныста. Дыбыстың әрең-әрең естілуіне сай келетін дыбыстық қысым шамасын *"естілу шегі"* деп атайды.

Егер дыбыс қысымын біртіндеп ұлғайтатын болсақ, дыбыстың қаттылығы артып, адамның есту мүшесін ауырта бастайды. Міне, осындай дыбыс қысымын, яғни *ауыртуды сезінудің шегі* деп атайды. Дыбыс қысымының *естілу шегі* мен *ауыртуды сезінудің шегі* арасындағы айырым мәнінің үлкендігінен дыбыс күшін

өлшеу үшін *лагорифмдік бірлік* қолданылады. Ол бірлік, белгілі физик Беллдың құрметіне – децибелл деп аталған (қысқаша – дБ). Қайсыбір шартты түрдегі нөлдік деңгей (стандарт) шамаларымен салыстыра қарағанда, аталған бірлік арқылы дыбыс күшін анықтауға болады. Мысалы, жиілігі *1000 Гц* болатын дыбыс үні үшін дыбыс күшінің деңгейі естілу шегінде – *0 дБ,* ал ауыртуды сезудің шегінде – *130 дБ.* Адамның есту мүшесінің сезімталдығы жиілік өзгерген сайын өзгеріп тұрады. Сондықтан дыбыстың қаттылығы оның жиілігіне тікелей байланысты. Акустика саласында кез келген дыбыс дауысының қаттылығы, осы дыбыс үнінің жиілік шамасы *1000 Гц* деңгейімен бірдей болатын дыбыс қысымымен өлшенеді. Аталған шама дыбыс қаттылығының деңгейі деп аталып, *децибелмен* өлшенеді. Әртүрлі жиіліктегі дыбыстар үшін бірдей дыбыс қаттылығына көрсететін қисық сызықтар *2-суретте* берілді.



2-сурет

Сурет бойынша дыбыс қаттылығы, жиілігі *1000 Гц-ке* тең болатын үннің дыбыс қаттылығымен бірдей болу үшін, дыбыстардың жоғарғылары *2000 Гц* жиіліктен жоғары және төменгілері *500 Гц* жиіліктен төмен болып келетін дыбыстардың интенсивтілігі басқаларға қарағанда әлдеқайда күштілеу болуы қажет.

3. Дыбыстың уақыттық сипаттамалары. Сөйлеу дыбыстарының сигналдарына олардың дыбысталу уақыттарының әртүрлі болу құбылысы тән. Бұл құбылыстың сегменттік элементтерінің артикуляторлық қасиеттеріне және барлық хабарламаның уақыт бойында таралуына қатысы бар. Соңғы аталғанға уақыттық айырмашылықтың әртүрлі қызмет атқаратыны белгілі. Айтылым кезінде сөйлеу дыбыстарының созылыңқылығы мен олардың тізбектері секундтың мыңдық үлестерімен, яғни миллисекундтермен (қысқаша – *мс*) өлшенеді. Созылыңқылықтың өзгеру шегі аса үлкен. Мысалы, еріндік-еріндік тоғысыңқы қатаң дауыссыз дыбыстың [Р] созылыңқылығы – *15 мс*, екпінді дауысты дыбыстың созылыңқылығы – *150 мс* пен *200 мс* аралығы, ал айтылымның созылыңқылығы ешбір фонетикалық шекпен шектелмейді, тек дыбысталуға тиісті мәтін ұзындығына ғана тәуелді болып келеді.

4. Дыбыстың сапасы және спектрлік сипаттамалары. Дыбыстардың акустикалық қасиеттері олардың сөйлеу жолында қалай жасалуына байланысты. Сөйлеу жолы дыбыстың қайнар көзі мен сөйлеу жолы қуысының көмейлік резонанстық қасиеттерінің әрекеттестіктерін қамтамасыз етеді. Сөйлеу жолының қайнар көзі – дауыс шымылдығының (желбезектің) тербелмелі қозғалысы негізінде пайда болатын төменгі жиілікті квазипериодтық үйлесімділік тербелмелі процесс. Дыбыстың бұл қайнар көзі дауыстылардың сонанттарды, ұяң шуыл дауыссыз дыбыстардың жасалуына қатынасады. Тоғысыңқы дауыссыз дыбыстардың жасалуының қайнар көзі ретінде дыбыстардың қабысу жеріндегі ауа қысымы мен атмосфералық қысымның теңесу процесі алынады. Осындай қысымдардың теңесуіне **импульстік шуыл** (шу) деп аталатын дыбыс пайда болады. Бұл дыбыс барлық жиілік диапозонында өзінің аз мөлшердегі созылыңқылығымен және тұрақсыз (жүйесіз) шуылдық сипатымен ерекшеленеді. Дыбыс саңылауы арқылы (щелевые согласные) жасалатын дауыссыздар ретінде шуыл қайнар көздің **турбуленттік** деп аталатын екінші түрі іске қосылады. Дыбыс қайнар көзінің бұл түрі сөйлеу жолы арқылы өтетін ауа ағымының үйкелісінің негізінде пайда болады. Яғни оларды дауыссыздардың артикуляция кезіндегі ауаның дыбыс саңылауынан үйкеліп шығуынан пайда болатын дыбысы деуге болады.

Сөйлеу жолының резонанстық жүйесі көмей қуысы арқылы жасалады. Оған жататындар: ауыз қуысы, жұтқыншақ қуысы және мұрын жолының қуысы. Ал мұрын жолы қуысының резонанстық жүйеге қосылуы тек қана ауа ағымына, оның жолы бос болған жағдайда ғана іске асады. Ауа ағымы үшін мұрын жолы қуысының босауы жұмсақ таңдайдың төменгі қалпында немесе оның жабық болуы – жұмсақ таңдайдың жоғары көтерілген қалпына сай келеді. Ауыз қуысы мен жұтқыншақ қуысы өзінің пішін үйлесімділігін, мөлшерін, ауа шығу тесігінің өлшемін тіл мен ерін қимылдарына қарай өзгертіп тұрады. Осы өзгерістер арқылы сөйлеу трактының резонанстық қасиеті анықталады. Сондықтан нақты дыбыстың артикуляциясында дыбыстың қайнар көзі – резонанстық жүйеде тербелмелі қозғалыс тудырады. Мұндай қозғалыстың қасиеттері артикуляция сипатына тікелей байланысты болып келеді.

Акустикалық тұрғыдан қарастырғанда (бір немесе бірнеше) дыбыс қайнар көздерінің резонанстық жиіліктерінің өзара әсерінің негізінде күрделі дыбыс туындайды. Мұндай дыбыстарды қабылдау мүмкіндігі олардың бірнеше акустикалық параметрлеріне байланысты болады.

4.4 Акустикалық сипаттамаларды талдаудың негізгі әдісі

Сөйлеу дыбыстарының артикуляциясы кезінде пайда болатын акустикалық сигналдардың ерекшелігі ретінде дыбыс қысымының өзгеруін сипаттайтын шаманың өте аз мөлшерде болуын айтуға болады. Осыған байланысты сөйлеу әрекетінің дыбысталу кезіндегі тербелмелі қозғалысты тікелей тіркеу мүмкін емес. Себебі, жоғарыда айтылғандай, зерттеушілер, бірінші реттегі сигналдардан гөрі, екінші реттегілерін көбірек қарастырады.

Акустикалық құбылыстарды тіркеудің негізгі екі түрі бар. Біріншісі, ол – арнайы аспаптар арқылы дыбыс толқынын толық қамтып алатын **осциллографтар**

деп аталатын сызба пішінін тіркеу. Екіншісі – алдын ала жүргізілген талдау негізінде қол жететін акустикалық параметрлерді тіркеу. Мұндай талдау жұмысы **спектроанализатр** деп аталатын арнайы аспап арқылы бірнеше тәсілдермен жүргізіледі. Дыбысты спектрлік талдау кезінде нақты дыбыстың туындауына қандай жиіліктер қатысады және олардың бір-біріне қатысты күші (интенсивтілігі) қандай деген мәліметтерге ие боламыз.

Мұндай талдау жүргізу үстінде сүзгілеу (фильтрация) әдісі қолданылады: екінші реттегі сигнал фильтрдің елеуішіне түседі немесе жиілігі өзгеріп отыратын фильтр арқылы өтеді. Осындай жағдайда, сүзгінің шығу тетігіндегі дыбыстың тербелісі фильтрдің жиілігінің тербелісімен сәйкес келеді. Мұндай тербеліс теңдіктерінен пайда болатын дыбыс қарқындылығы талданып жатқан дыбысқа сәйкес жиілік құрастырушысының интенсивтілігіне тікелей қатысты болып келеді. Талдау нәтижесінде амплитудты-жиілік спектрлері немесе спектрдің лездік шамасы анықталады. Егер бір дыбыстың бірінен кейін бірі жалғасып келетін қысқа бөліктерінің әрбіреуіне қатысты спектрлердің аталған лездік шамаларын анықтайтын болсақ, онда мұндай талдау дыбыс спектрінің уақытқа қатысты өзгерісі жайлы түсінік береді.

Дыбыстардың уақытқа қатысты спектрлік сипаттамаларын бұдан да тез талдау үшін **динамикалық спектрограф** деп аталатын арнайы аспап құралы пайдаланылады. **Динамикалық спектрограммада** жиілік осі тік (вертикаль) орналасады, ал көлденең ось бойынша уақыт шамасы салынады. Дыбыстың жиіліктік құрастырушыларының интенсивтік мәні сәуле түсіру нүктелерінің анықтығына қатысты болады. **Спектрограммаларды** жарық сезгіш таспаға немесе қағазға түсіретін болсақ, сәуленің жарықтығының ұлғаю сипатына қарай, олар қараңғылау болып, қарауытып түседі. Мұндай жағдайда спектр құрастырушыларының салыстырмалы интенсивтілігін адам көзбен қарап-ақ ажыратуына мүмкіндігі бар. Күні кеше ғана сөйлеу тілін **осциллографтық** және **спектрлік** талдаулар **аналогтық жүйелер** деп аталатын әдіспен жүргізіліп келген. Соңғы жылдары, есептеу техникасының дамуының нәтижесінде жаппай түрғыда **сандық талдау әдісі** қолдау тауып, дыбысты акустикалық тұрғыдан талдау айтарлықтай ұтымды және сапалы бола бастады.

Сөйлеу тілінің сигналдарын талдау үшін арнайы құрастырылған **аналогты-сандық түрлендіргіштер** іске қосылды. Олардың көмегімен әрбір дыбысты немесе дыбыстық тізбектерді бүтін сандар жиынтығы ретінде түсінік беруге және сандар жиынтығы түрінде компьютер жадына жазып алуға мүмкіндік туады. Арнайы жазылған компьютерлік бағдарламалар арқылы дыбыстың тиісті акустикалық параметрлерін тіркеп және оларға талдау жүргізіп, бұрынғы осциллографтар мен спектрографтар көмегімен жүргізілген талдау нәтижелерінің берілу пішінінде ұсынуға мүмкіндік туады. Сол сияқты, бағдарламалар арқылы компьютер жадына жазылып қойылған дыбыстардың жиіліктік құрастырушыларын, қажеттікке сай не күшейтуге, не әлсіретуге, не өте қысқа кесінділер түрінде сегменттеуге және т.б. іске асыруға қол жетті деуге болады. Мұндай нәтижелер дыбыстардың акустикалық қасиеттерінің маңызды жақтарын зерттеп-білуде және оларды қабылдауда аса маңызды деп саналады. Дыбыстарды бұлайша талдауды **сандық-аналогтық** деп аталатын **түрлендіргіш** арқылы да іске асыруға болады. Бұл қондырғының көмегімен сандар жиынтығы пішініндегі дыбыс **аналогтық акустикалық сигнал** түрінде қайта өндіріледі.

Сөйлеу сигналының акустикалық қасиеттерін талдаудың арнайы түрі ретінде дыбыстау мүшесінің жиілігінің өзгеру сипатын талдау болып саналады. Мұндай талдау **интонация** мәселесін зерттеуде аса қажет. Себебі **интонацияның** негізгі компоненті – **әуен** (мелодика) – дыбыс желбезегі арқылы жасалатын тербелмелі қозғалысының өзгеру сипатына тікелей байланысты болып келеді. **Осциллографтық қисық сызық негізгі тон периодының шамасын** өлшеу үшін жеткілікті мәліметтер береді және **интонация** мәселесіне арналған көптеген **эксперименттi-фонетикалық** зерттеулер осындай өлшемдерге негізделеді. Осциллограммамен жұмыс істерде әрбір период шамасы өлшенеді немесе бір топ кезеңнің созылыңқылығы анықталады.

Әрине, бірінші жағдайда әрбір кезеңнің шамасын өлшеу оңайға түспейді, ал екіншіде – зерттеушіні қызықтыратын өзгерістер жайлы орташа мәлімет алынады. Сол сияқты, дыбыс желбезегінің тербелу периодын өлшеу мен жекелеп бөліп алуды автоматты түрде іске асыратын арнайы аспаптардың бар екендігін айтуға болады. Мұндай талдау, қазіргі кезде, арнайы бағдарламалар (программалар) жазу негізінде компьютер көмегімен жүргізіледі.

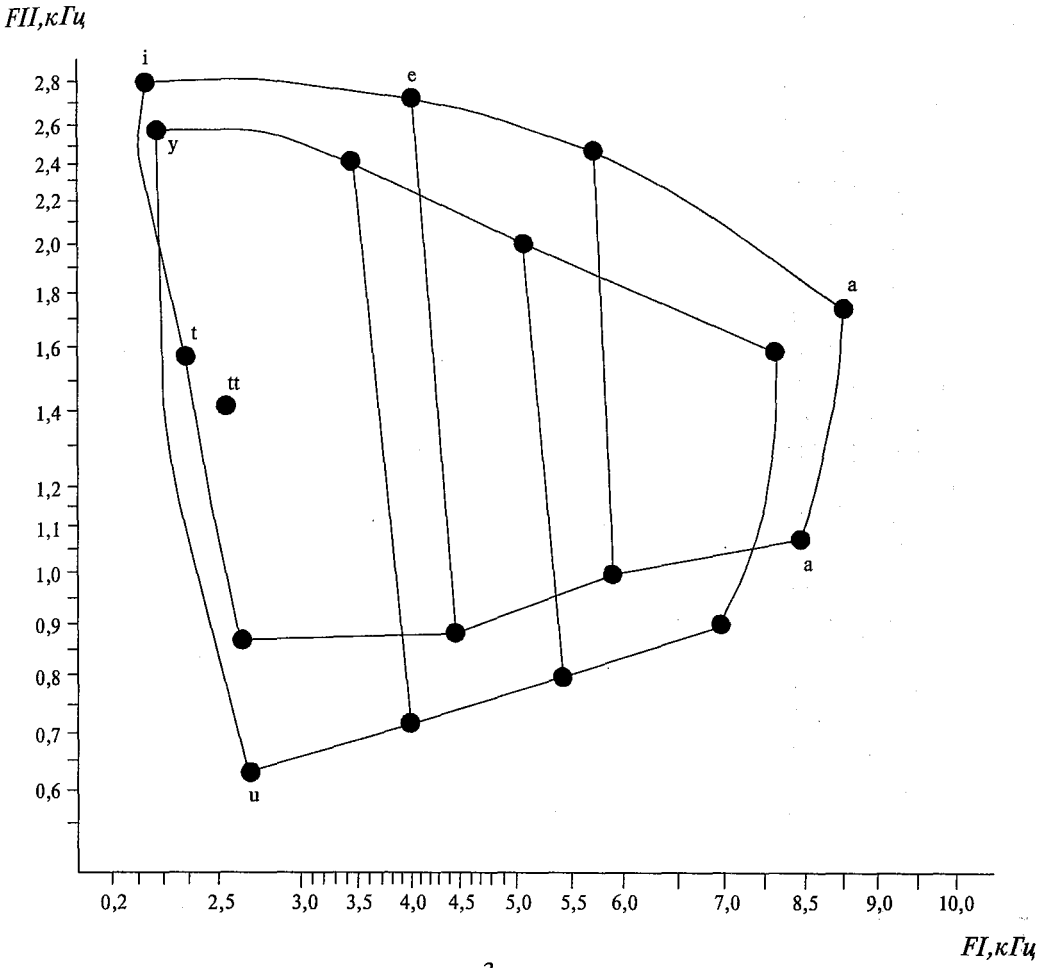
4.5. Динамикалық спектрограммалардағы дауысты және дауыссыз дыбыстар

Қолданбалы фонетикада динамикалық спектрограммалар кең қолданыс тауып жүр. Мұндай спектрограммаларға дұрыс түсінік беру үшін, олардың қасиеттеріне Л.В.Бондарконың "Фонетические аспекты прикладной лингвистики" атты мақаласындағы орыс тілінің дауысты және дауыссыз дыбыстары негізінде берілген мәліметтеріне толығырақ тоқталайық [Прикладное языкознание, 1996, 121-123-бб.].

Дауыстылар арасындағы айырмашылықтар оларды дыбыстау кезіндегі көмейлік резонаторлардың айырым белгілері арқылы анықталады. Резонатор тәрізді рөл атқаратын сөйлеу жолының (трактының) қуыстарында тербелмелі процестердің қозуы байқалады. Осының әсерінен акустикалық тұрғыда қарастырғанда күрделі дыбыс пайда болады. Ондай дыбысты сипаттау үшін олардағы қай спектрлер аймағының күшейтілгендігін анықтау қажет.

Сөйлеу жолының резонанстық сипаттамаларына байланысты болатын дауысты дыбыстардың спектріндегі күшейген аймақтарды *форманттар* деп атап, оны латынның ***F*** әрпімен белгілейді. Дауыстыны жан-жақты сипаттау үшін, яғни ондай сипаттамалар бойынша нақты бір дауыстыны басқаларынан айырып алу үшін алғашқы ***FI*** мен ***FII*** форманттардың жиіліктері ескерілсе жеткілікті деп ұйғарады. Жоғары жиілікті ***FII*** дауыстылар үшін ***FIII*** сипаттамасы да маңызды болып саналады.

Дауыстылардың форманттық жиіліктері мен артикуляциялық белгілері арасында тәуелділік бар. Бірінші формант ***(FI)*** дауыстының көтерілу дәрежесінің белгілерімен қатынаста болады. Мәселен, дауысты дыбыс қаншалықты ашық болған сайын, бірінші форманттың жиілігі жоғары болады. Екінші формант бірнеше дауыстылармен байланысты: ең алдыңғы дауысты екінші форманттың ең жоғары жиілігіне, ал ең артқы – барынша төмен жиілікке ие болады. Дауыстылардың еріндіктері барлық форманттарының жиіліктерін төмендетеді. Осындай жалпы фонетикалық тәуелділік *3-суретте* көрініс тапты. Онда дауысты дыбыстар форманттарының жиіліктері мен артикуляциялық белгілері арасындағы қатынасты анықтайтын диаграмма берілген.



3-сурет

Динамикалық спектрограммада әрбір дауысты дыбыс үнемі өзгерістегі акустикалық көрініс түрінде орын алады. Яғни ең жақсы деген жағдайда, дауыстының форманттық жиілігінің тұрақты бөлігінің алдында бірінші өтпелі бөлік тұрады. Бұл бөліктің қасиеттері алдындағы дыбыстың сипаттамаларына тәуелді болып келеді. Тұрақты бөліктен кейін келесі дауыссызға тәуелді екінші өтпелі бөлік те кездеседі. Кейбір жағдайда аталған бөліктер ішінен бір немесе екі бөлігі жоқ болып шығуы да мүмкін. Барлық дауыстыларды сипаттайтын барынша жалпылаушы белгі, ол айқын форманттық құрылым мен дыбыстың жоғары интенсивтілігі. Бұл екі қасиет те дыбыс көзінің болуымен және резонанстық жүйенің активтік рөлімен анықталады.

Дауыссыздардың динамикалық спектрограммалары олардың жасалуына қатынасатын қайнар көздің (дауыс көзі, импульстік көз, турбуленттік көз) қасиеттері жайлы ақпаратқа ие және сөйлеу жолындағы (дауыссыздардың) орнын да анықтайды.

Сонымен, жоғарыда қысқаша сөз болған дауысты мен дауыссыздардың негізгі акустикалық қасиеттері сөйлеу тілін автоматты түрде танып-білуге аса қажет деп саналады.

4.6 Тілді автоматты танудың фонетикалық мәселелері

Тілді автоматты түрде тану проблемасында танымдық жүйенің мынадай қасиеттері маңызды болып саналады:

- танымға қажетті тілдік бірліктердің саны;
- диктордың дауысына байланысты шектеулер;
- танымға жататын хабардың қасиеттері;
- танымның уақытқа қатыстылығы;
- танып-білудің шарты.

Осы аталғандардың әрбіреуіне арнайы тоқталайық.

1. Танымға қажетті тілдік бірліктердің саны. Техника саласында шешуді қажет ететін негізгі мәселелердің бірі – көлемі шағын бұйрықтарды немесе сандарды танып-білу мәселесі. Мұндай есептерді шешу үшін сөздің толық пішінді акустикалық бейнесін талдау қажет. Сонымен бірге, таным процесінің берік және тиімді белгілерін айқындап алу қажет болады. Мысалы, *"1"* саны мен *"9"* санының және олардың аралығында жатқан сандардың атауларын танып-білу үшін олардағы буындар саны немесе атау сөздерінің соңғы дыбысының дауысты не дауыссыз болып келуі арнайы белгілер ретінде алынуы мүмкін. Бұл белгілер тек осы сөздер топтамаларына ғана тән және танымдық аспаптың жадынан көп орын алмайды. Ал егер тануға тиісті сөздер топтамасы сан жағынан көп болатын болса, жоғарыда аталған әдіс қолайсыздық тудыруы мүмкін. Мұндай жағдайда басқаша, яғни әр сөзді кішігірім дыбыс сегменттері ретінде қарастырған тиімді болады. Фонетист маман үшін мұндай дыбыстық бірліктердің санын анықтау маңызды мәселе деуге болады. Бұл жердегі ең бір басы ашық нәрсе – ондай санның барлық фонемалар санынан аз болуы. Бірақ көптеген танымдық жүйелерде қолданыс табатын дыбыстық бірліктер сегменттерінің саны фонемалар мөлшерінен көп болып кездеседі. Мұндай дыбыстық бірліктер *"тану фонемалары"* деп аталып жүр. Тану кезіндегі қолданылатын бірліктердің санына қарай, олардың айырымдық сипаты негізінде қажетті деген белгілерді таңдап алу іске асады.

2. Диктордың дауысына байланысты шектеулер (дикторды таңдау). Танымдық жүйе бір адамның ғана дауысымен дыбысталған бұйрықтарды анықтау үшін арналған болуы мүмкін. Бұл ең жеңіл жол, себебі сөйлеу сигналының акустикалық қасиеттерінің түр-түрі дыбыс сапасымен тығыз байланысты болып келеді.

Егер тану жүйесіне кез келген дыбысты танып-білу қажет болса, тану мәселесі қиындайды, себебі ондай аспап не әрбір нақты диктордың дауысына икемделіп отыруы керек немесе барлық дикторлардың дауыс ерекшеліктеріне ортақ сипатқа икемделуі керек. Осыған байланысты дауыстың жекелік ерекшеліктерінің сөйлеу сигналдарының акустикалық қасиеттеріне әсер ететіндігін ескеруден келіп, мынадай ой туындайды – дауыстың акустикалық көрінісіне қарай, айтылымның нақты дикторға қатынастылығын анықтау.

3. Танымға жататын хабардың қасиеттері. Күрделі танымдық жүйелері арқылы қиындасқан мәтіндерді танып-білу мүмкіндіктерін қарастырады. Бұл мәселенің шешімін табу, қомақты сөздіктерді танып-білуден қиындау. Сөйлеу түріндегі хабарламаны құрайтын дыбыс элементтерінің акустикалық қасиеттері сегменттік және суперсегменттік сипаттамалардың өзара әрекеттенуінен көптеген өзгерістерге ұшырайды. Сонымен бірге, кез келген айтылымға тән мағыналық артықтықтың нәтижесінде фонетикалық белгісіздік те арта түседі. Бұл дәстүрлі жағдайдағы фонетиканың өзіне ғана тән сипаттамалар бірізділігіне күмән туғызуы

мүмкін. Осындай танымдық жүйелердің даму процесінен байқайтынымыз, шешілетін мәселелер күрделенілген сайын, қосымша ақпарат үлесі де өз тарапынан арта түседі. Мұндай біріздікті талап ететін қосымша ақпараттар да лингвистикалық сипатта болатыны белгілі. Қиындасқан мәтінді танып-білу кезіндегі тану жүйесіне қатысты қосымша ақпаратқа, мысалы, орыс тіліндегі мәтіндер үшін, фонетикалық тұрғыдан анықталмаған дауысты дыбыстардың фонемдік түсінігінің ережелерін немесе сөзтұлғаның екпін түсетін бөлігіндегі фонетикалық пен фонемдік сегменттелуінің арақатынасын жатқызуға болар еді. Ал танымдық жүйеге қосымша енгізілетін шектеулер ретінде сөйлемдердегі сөздердің кездесу ретін немесе сөйлеу қарқынына (темпіне) қойылатын шектеулерді мысал ретінде келтіруге болады.

Нақты тілдің көлем жағынан алғанда шексіз сипаттағы сөздігін танып-білуге арналған танымдық жүйелерінің жұмысы күрделі лингвистикалық ережелерге негізделген болуы керек. Ондай ережелер сөйлеу түріндегі хабарламаның тек акустикалық құбылу сипатын ғана ескеріп қоймай, сонымен бірге дыбыс тізбегінің статистикалық қасиеттерін және айтылымның фонетикалық толықтығының жасалу зандылықтарын да есепке алуы қажет.

4. Танымның уақытқа қатыстылығы. Табиғи тілдік қатынаста хабарлама мазмұнын танып-білу уақыттың ақиқат өлшемінде орындалады. Сондықтан танудың автоматты жүйесін құрушылар акустикалық сигналды тануды да табиғи жағдайға жақындатуға тырысады. Бұл үшін ең алдымен өте маңызды екі мәселені шешіп алу қажет. Олар, біріншіден, акустикалық жағынан үздіксіз болып келетін айтылымда "танымдық фонемамен" сай келетін минимальді бірліктерге сегменттеу. Екіншіден, ондай сегменттерді танымдық аспаптың жадында сақталған эталондық үлгілермен сәйкестендіру мәселесін шешу. Аталған екі мәселенің шешімін табу, сөйлеу технологиясымен айналысатын мамандар үшін алғашында қарапайым болып көрінгенімен, олар шындығына келгенде басқаша – күрделі мәселе. Себебі, дыбыстық сипаттағы сөйлеу процесімен айналысу кезінде адам баласы ондай сегменттеу мен сәйкестендіру мәселелерінің шешімін табуда акустикалық емес, лингвистикалық негізде іске асырады. Мұндай жағдай сөйлеу процесін автоматты тануды екі түрлі жолға сілтеуі мүмкін. Біріншісі – адамның сөйлеу қызметінің моделін құру жолында аталған екі мәселе де лингвистикалық ережелерге жүгінуді қажет етеді. Яғни, тіпті сөйлеудің акустикалық көрінісінен алдын ала келісілген мәтіндік "транскрипцияға" көшірудің өзінде-ақ күрделі лингвистикалық амалдар қажет. Екіншісі – адамның сөйлеу қызметін модельдеуден бас тарта отырып, тек танымдық аспаптың мүмкіндігіне ғана бейімделген акустикалық сигналдар түсінігін сипаттай алатын алгоритмдер құру.

5. Танып-білудің шарттары. Байланыс арнасымен сөйлеу тіліндегі ақпараттарды тарату әртүрлі жағдайларға қатысты болуы мүмкін. Олардың көпшілігінде таралу арнасынан естілетін шулы дыбыстар ішінен сөйлеу сигналын бөліп алуға тура келеді. Бұл әрекет автоматтанған таным процесіне де тән деуге болады. Мұндай жағдай зерттеу аясындағы сигналдардың акустикалық сипаттамалары бойынша "сөйлеу – сөйлеу емес" ("речь – не речь") деп аталатын жаңа ғылыми бағытты өмірге әкелді. Шулы дыбыстар ішінен сөйлеуге тән сигналды танып-білу адамның сөйлеу қызметіндегі дәл осы процесті модельдеу мүмкіндігін айқындауды талап етеді. Сондықтан аса шулы жағдайдағы нақты дыбыстың белгілерін айқындау автоматтанған таным процесінің аса бір қиын мәселесі болып отыр.

4.7 Зерттеу мен қолданбалы бағдарламалар арқылы сөйлеу әрекетін синтездеу мәселесі

Адамның сөйлеу әрекеті кезінде пайда болатын дыбыстарға ұқсас келетін дыбыстарды туындату мәселесі ғалымдарды бұрыннан-ақ қызықтырып келгені белгілі. Алғашында мұндай дыбыстарды туындату таза рефлекстік сипатта болса да, оның тілдік бірліктердің фонетикалық белгілерін анықтай білуде аса зор рөлі болды.

Қазіргі кездегі іске қосылған синтездеуші аспаптарды құрастыруға қатысты алға қойған мақсатқа сай және синтезделінген сөйлеу тілін алу әдіс-тәсілдеріне қарай, оларды топ-топқа бөліп қарастыруға болады.

Адамның ақпаратты қабылдау механизмін зерттеу аясында синтезделінген сөйлеу тілі кең қолданыс тауып жүр. Зерттеулер нәтижелері адам қоғамындағы тілдік қарым-қатынасты жетілдіруге әсері бар психофизиологиялық процестерді модельдеуде кеңінен қолданылуда. Осындай тәжірибелерде қолданылатын синтезделген сигналдардың құндылығы мынада. Біріншіден, зерттеуші өз болжамына қарай, синтезделген сигналдар сипаттамаларын өзгерте алу мүмкіндігі бар. Екіншіден, синтезделген сигналдың қабылдаушыға тигізетін әсерін тексеруге де мүмкіндік туындайды.

Синтезделген сөйлеу тілінің практикалық қолданысы адамдардың әртүрлі техникалық құрылғылармен табиғи тілде тілдесуінің аса ыңғайлы болу жағдайын көздейді. Мысалы, оған соңғы жылдары синтезделген сөйлеу тілін автомобиль жүргізуді компьютермен басқаруда қолданып, оны табиғи тілде іске асыратын жүйесі жайлы сөз етуге болады. Автомобильдің қозғалысы кезінде бас компьютер жолдың сапасы жайлы және оған сай қандай жылдамдықты ұстаным етуін, жанармай мөлшерін, салқындатылатын жүйедегі температуралық режимді және т.б. хабардар етіп отырады. Сол сияқты, егер кабина мен жүксалғыш есігін жүргізуші кезінде дұрыс жаппай кетсе, олар дұрыс қалыпқа келмейінше компьютер әрі қарайға қозғалысқа рұқсат бермеуі мүмкін (автомобиль от алмай қояды).

Жазба мәтін бойынша да сөйлеу тілі сигналына синтез жасау ғылыми бағыты көптеп қолданыс табуда. Мұндай жүйелер қоғамдық мақсаттар үшін, яғни көру мүмкіндіктері нашарлаған немесе одан толығымен айырылған адамдар үшін өте тиімді қолданылып жүр. "Мәтін – сөйлеу тілі" деп аталатын синтездеу жүйесі кез келген мәтінді дыбыстата алады. Басқа сөзбен айтқанда, бұл жүйе жұмысының нәтижесі көру қабілетінен айырылған адамдардың мәтіндерді "оқуына" және осы әдіс негізінде ондай кемтар адамдарды оқытуға мүмкіндік туғызып отыр.

Мұндай жүйелерді құруда қайсыбір аралық мәселеге қатысты жайттардың шешімін табу жағдайы туындайды. Олар әрі фонетикалық, әрі техникалық факторларға байланысты – орфографиялық мәтінді фонетикалық символдардың тізбегіне ауыстыра алатын транскрипциялау проблемасы. Мұндай күрделі мәселенің шешімін табу үшін арнайы компьютерлік бағдарламалардың көмегімен кез келген мәтінді фонетикалық пішінге келтіре алатын автоматтанған транскрипция жасау жүйесін құру қажет. Мысалы, орыс тілі үшін жазылған мұндай бағдарламалар, мәтіндегі әрбір сөзге екпін орнын алдын ала көрсетіп беруді талап етеді. Әрине, бұл мәселені толық автоматтандыру үшін, ең алдымен алдын ала мәтінді қолмен өңдеуді қажет етпейтін, лингвистикалық ережелердің аса күрделі түрін қолдану керек.

Сапалы түрдегі синтезделінген, яғни табиғи дыбысталатын, анық естілетін сөйлеу тілін жасау үшін таңбалардың фонетикалық транскрипциялары мен оларға

сәйкес келетін аллофондардың акустикалық өлшемдерінің арақатынасын анықтау қажет. Осындай мәліметтерді қорландыру – қолданбалы фонетиканың маңызды мәселелерінің бірі. Жоғары сапалы синтез жасаудың тағы бір шарты – айтылымның интонациялық безендірілу ережелерін алгоритмдеу. Бұл мәселенің шешімінің табылуы қиындасқан сөйлеу тілінің интонациялық аясында зерттеулер жүргізуге ниеттендіреді.

Енді адамның қатысынсыз, басқа да қайнар көздер арқылы пайда болатын сигналдардың сөйлеу тіліне ұқсас болып келетін синтезделінген түрлерін алу жолдарын қарастырайық.

Шартты түрде олар екі топқа бөлінеді: **техникалық синтез** және **компилятивті синтез.**

Техникалық синтезде сөйлеу тіліне ұқсас сигналдарды туындату техникалық құрылғы көмегімен немесе арнайы компьютерлік бағдарламалар арқылы орындалады. Ал **компилятивті синтезде** сөздер, фразалар немесе толық мәтін түрлері табиғи сөйлеу тілінің дыбыс тізбектерінің негізінде іске асады. **Техникалық синтезді** сөйлеу тілін туындатудың модельдік әдісі деп сипаттауға болады.

Синтездеу процесінің даму тарихы бірнеше кезеңдерге бөлінеді. Мәселен, XVIII ғ. соңында адамның сөйлеу жолына ұқсас, ал дауыс шымылдығы ретінде тербелмелі тілшік тәріздес келетін акустикалық резонаторға негізделген ең алғашқы "сөйлейтін" машина жасалған болатын. Бұдан кейін де осындай сөйлейтін машиналар түрлері көптеп орын алды. Олардың бәрінің де жұмыс істеу принципі адамның сөйлеу тілін туындатуына ұқсас және адам сөйлеу жолы құрылымын қайталайды. Бұл аталған техникалық синтездің даму жолын механикалық кезең деп атауға болады.

Электроакустикалық саланың дамуымен бірге синтездеу құрылғысын жасау жолы да жаңа бағытқа ие болды. Енді электр күшімен жұмыс істейтін синтезаторлар құрастырыла бастады. Негізінен мұндай жүйе жұмысы да сөйлеу кезіндегі дыбыс қуысының дыбыстық мүшелерінің артикуляторлық қозғалысын модельдеуге бағытталған болатын. Басқа сөзбен айтқанда, ондай жүйенің кейбір элементтері сөйлеу жолының белгілі бір бөліктерін қайталайды.

Қазіргі кездегі синтезаторлар да сөйлеу трактына ұқсас, бірақ олар негізінен, сигналдардың акустикалық қасиеттерін туындатудағы жол бөліктерінің рөлін анықтауға көп көңіл бөледі. Синтезаторлардың басқа бір түрі – дыбыстардың акустикалық қасиеттері бойынша сөйлеу тіліне ұқсас келетін жақтарын ғана синтездеуге негізделген.

Спектрлі-уақыттық көріністің жалпы жағдайы синтезделсе де немесе табиғи дыбыстардың қайталанатын қасиеттері синтезделсе де, акустикалық процестің уақытқа қатысты пайда болуында қайнар көз сипаттамалары ескерілуі қажет. Есептеу техникасының дамуы, электронды-есептеу машиналарды (ЭЕМ) синтездеу құралы ретінде пайдалануға мүмкіндік туғызды. ЭЕМ-нің көлемді ақпараттарды жадында сақтай алуы және логикалық амалдарды аса тез жылдамдықпен орындай алатын мүмкіндіктері бар екендігі көпшілікке мәлім. Сондықтан ондай машиналарды аналогтық дыбыс жолдарын басқару ісіне де және форманттық синтездеу мен спектр синтезін алу ісіне де қолдануға болады.

Компилятивті синтез есептеу техникасының даму салдарынан, техникалық тұрғыда алғанда, ақиқаттық әдіс ретінде пайда болды. Қазіргі кездегі компьютерлер сөйлеу тілі сигналын цифрлық пішінге келтіре алады. Сонымен бірге олар сигналдарды ұзақ мерзімге жадында сақтауға, өңдеуге және әртүрлі талдаулар

жүргізуге мүмкіндігі мол. Зерттеуші компьютер көмегімен фонетикалық транскрипцияланған таңбалардан ақиқат дыбыстауға өтуге қажетті табиғи сегменттердің жиынтығын да ала алады.

Компилятивті синтездің жүйесін құрудың негізгі мақсаты – өзара сәйкестікте болатын кішігірім элементтер жиынын жасау. Осымен бірге **компилятивті синтез** арқылы табиғи просодиканы құрастыру мәселесі де маңызды болып саналады.

Алғашқы тәжірибелерге қарағанда **компилятивті синтез** әртүрлі созылыңқылық сипаттағы элементтерге негізделеді, яғни негіз сөзге сәйкес келетін дыбыстар тізбегінен дауысты бөліктің ауыспалы шегіне дейінгі аралыққа негізделеді. Қазіргі кезде **компилятивті синтез** шет тілдерін үйрету бағдарламаларында және көру мүмкіндігі бұзылған адамдарды оқытуда кең қолданыс тауып жүр.

4.8 Артикуляцияны зерттеудің негізгі мәселелері

Адамдардың сөйлеу әрекеті аса күрделі процеске жататыны жайлы жоғарыда да айтқан болатынбыз. Дыбыс шығарудағы сөйлеу мүшелерінің тікелей байқалатын әрекеттері аса күрделі және рет-ретімен орындалатын әрекеттердің соңғы нәтижесі деуге болады. Дыбыстау мүшелері көмегімен және біртіндеп орындалатын бұйрықтар негізінде, айтылуға тиісті хабардың алдын ала белгілі семантикасының мазмұны ашыла бастайды. Артикуляциялық процесс туралы дұрыс түсінік алу сөйлеу тілінің синтезін жасау кезінде қажет екені белгілі.

Сөйлеу тілін автоматты тану процесі күрделі мәселе және өзіне тән қиындықтары да бар. Осындай қиындықтардың шешімін табуда артикуляция мәселесіне арналған зерттеу жұмыстарының құндылығы бұрынғыдан да арта түседі.

Өткен ғасырдың 60-жылдарында "адамның моторлы орталықтарының қызметі акустикалық сигналдармен жан-жақты әрекет жасауға мүмкіндік туғызады" деген тұжырым пайда болды.

Сонымен қатар мынадай ұйғарым жасаушылар да болды: физиологиялық және фонетикалық көзқараспен қарастырғанда, фонемалар мағына ажырату элементі ретінде тек қана артикуляторлық программалар арқылы синтездеу деңгейінде ғана қолданылады.

Артикуляциялық процестің құрылымын түсінуге талпыныстан барып, артикуляцияны талдаудың жаңа әдістері пайда болды. Ал оларға қойылатын негізгі шарт – зерттеу мақсатына сай келуі.

4.9 Артикуляцияны зерттеу әдістері

Эксперименттік фонетика сөйлеу тілі дыбысының физиологиясын зерттеу ғылымы деген атпен жұртшылыққа аян. Оның алғашқы әдістемелік бағыты дыбыстың ең маңызды деген артикуляциялық белгілерін тіркеу мынадай мәселелерге арналды:

- сөйлеу кезіндегі дем алу мәселесі (речевое дыхание);
- дыбыстау мүшелерінің орналасу қалпын айқындау;
- дауыс шымылдығының әрекеттік күйі (поведение голосовых связок) және т.б.

Ертеректегі зерттеу жұмыстарында тәжірибеші әртүрлі дыбыстарды дыбыстау кезіндегі өз бұлшық еттерінің сезінулерін талдау арқылы, яғни өзін-өзі байқау негізінде артикуляция қасиеттерін анықтауға талпынған болатын. Әдетте, алғашқы

байқаулар дыбыс мүшелерінің орналасу қалпын сипаттауға бағытталды. Бұл зерттеу бағыты – статикалық бағыт.

Уақытқа қатысты артикуляцияның өрістеуі (развертывание) жайлы түсінік артикуляцияның үш фазасына байланысты ілімімен шектеледі. Олар – **бастапқы** (приступ), **негізгі бөлік** (выдержка) және **аяқты бөлік** (отступ). Бұл үшеуінің ішінен екіншісі, яғни **негізгі бөлік** (выдержка) фазасы аса маңызды саналады.

Физиологиялық фонетиканың алғашқы кезеңіндегі аспаптар арқылы артикуляцияны талдау, еріндік артикуляцияны фотосуретке түсіруге, палатография мен рентгенография нәтижелеріне негізделген болатын. Енді осы аталғандарға қысқаша тоқталайық.

Еріндік артикуляцияны фотосуретке түсіру артикуляциялық белгіні тіркеудің, яғни дауысты дыбыстар мен кейбір дауыссыз дыбыстардағы ерін қозғалысын (не тұрақты қалпын) белгілеудегі ең ұтымды тәсілі деуге болады. Мұндай мәліметтерді жете талдаудың негізінде мынадай жайттар анықталды. Мәселен, еріндік артикуляция арқылы тек еріндік емес, басқа да дыбыстарды танып білуге болатындығы. Оған дәлел болатын мысал – есту қабілетіне жартылай не толық зақым келгендерді ерін қимылдары арқылы тілдесу мүмкіндігі. Сонымен бірге бұл әдіс дыбыссыз тілдесу қажеттігі үшін де қолданылады.

Палатография – ол артикуляция кезінде тілдің қатты таңдаймен түйісетін жерін (орнын) анықтау үшін қолданатын әдіс. Ол үшін сынаққа түскен адамның таңдайының пішінін дәлме-дәл қайталайтын жұқа табақшадан (пластинкадан) жасанды қатты таңдай дайындалады. Жасанды таңдаймен жұмыс істеу реті былайша өтеді. Ең алдымен, таңдайдың рөлін атқаратын тез алып тасталынатын табақшаны арнайы маймен (немесе талқа ұнтағымен) майлап, оны сынаққа қатысушы адамның ауыз қуысына апарады. Қажетті деген дыбыс айтылғаннан кейін, табақша ақырын алынып, жасанды таңдайдағы таңбаланып қалған тіл ізінің суретін қағазға не фотоға түсіріп алады. Қазіргі кезде бұл әдіс барынша жеңіл. Яғни жасанды таңдайды қолданбай, оның орнына адам таңдайына бояу жағылады да тиісті дыбыстың артикуляциясы аяқталғаннан кейін, ауыз қуысында орналасқан арнайы айнадағы шағылысуы фотосуретке түсіріледі.

Артикуляция кезіндегі тіл қалпы жайлы ақпарат алуды рентгенография әдісі арқылы жүргізеді. Осы әдіске қысқаша тоқталайық.

Ең алдымен, тілдің орта сызығының бойымен, рентген сәулесін өткізбейтін арнайы қоспа бояумен сызық жүргізіледі. Осы сызық арқылы рентгенограмма жәрдемімен айтылған дыбысқа қатысты тілдің пішін үйлесімділігін байқауға болады.

Экспериментті фонетиканың әрі қарайғы тәжірибелерінде **кинорентгенографтау** әдісі қолданыла бастады. Бұл әдіс бойынша артикуляторлық процестің динамикасы жайлы мәліметтерді де алуға болады. Бірақ қазіргі кезде аталған екі әдіс те айтарлықтай кең қолданыс таппай жүр. Оның себебі тәжірибеге қатысушы адамға ұзақ уақыт бойы рентген сәулесін қабылдаудың зияндығы, аспап-құралдардың күрделілігі және графикалық материалдарды өңдеудің қиындығы деуге болады.

Сөйлеу тілін туындатудың динамикасын зерттеу жұмыстарының ішінен ең нәтижелісі деп, ғалым-фонетист Л.Г.Скалозубтың зерттеулерін айрықша айтуға болады. Осы ғалымның еңбектерінен оқырман артикуляторлық қозғалысты жүйелі түрде талдау кезінде туындайтын көптеген қиындықтарға дәлел болатын фактілерді кездестіре алады.

Өткен ғасырдың 60-жылдарында сөйлеу тілінің артикуляторлық өлшемдерін тіркеудің, белгілі дәрежеде, бірегей әмбебап сипаттағы комплексті жүйесі құрылған

болатын. Бұл жүйе неше түрлі лезде хабар беретін (датчик) құрылғылардан және олардан таралатын сигналдарды күшейту мен алдын ала өңдеуге қатысты сызбалардан тұрады. Сонымен қатар аталған жүйе ақпаратты өңдейтін логикалық сызбалардан, тіркегіштерден және автоматты өлшеуіштерден (измерители) тұрады.

Электр сигналдары түрінде бір мезгілде және үздіксіз сипаттағы мынадай көрсеткіштерді алуға мүмкіндік бар: ішкі қысым, фонация кезінде пайда болатын ауа қозғалысының жылдамдығы, ауа түрткісі (толчки), іштің және көкіректің әр деңгейлеріндегі демалыс қозғалыстары, бет құрылымындағы төменгі жақтың қозғалысы, еріндердің түйісуі мен ажырауы, еріндік құбылыс, динамикалық палатограмма. Осы аталған артикуляторлық параметрлерді тіркеумен бірге мұндай жүйе артикуляцияның нәтижесі ретіндегі әдеттегі дыбыстық жүйені жазып алуға да мүмкіндік туғызады.

Айтылым элементтерін артикуляторлық сипатта құру. Артикуляцияны талдаудың жетілген әдістерін пайдалану теориялық және практикалық фонетиканың маңызды мәселелеріне жаңаша қарауға мүмкіндік туғызды. Ең алдымен, нақты дыбыстардың айтылу динамикасына талдау жүргізуге болатын жағдай жасалды. Әрбір дыбыстың уақытқа қатысты бағдарламасының үшфазалық құрылымы жайлы классикалық көзқарас шындыққа сай келмейтіні анықталды. Үздіксіз дыбыстық тізбеде бірінші элементтің аяққы дыбысы (отступ) одан кейінгі келетін элементтің бастапқысымен сәйкес келеді, бірақ классикалық сызбадан ауытқу тек мұнымен шектелмейді. Табиғи тілдің сөйлеу ағымында **негізгі бөлік** фазасының (фаза выдержки) да түсіп қалуы жиі кездеседі. Бұл жағдай, әрбір бөлек дыбыстың дыбысталуы жалпы **айтылым бағдарламасына** қатыссыз деген болжам жасауға мәжбүр етеді. Арнайы жүргізілген тәжірибелердің нәтижесі бойынша, бір ғана **айтылым бағдарламасында** орындалатын ең кіші айтылым бірлігі ретінде **ашық буын** саналады.

Үздіксіз сөйлеу тілі тізбегінде артикуляторлық қозғалыстар ең қарапайым пішіндегі алгоритм жолымен құрылған: сөйлеу тізбегі – ол буындар тізбегі, ал әрбір буын алдыңғы дауыссыздардың саны мен сапасына тәуелсіз дауыстымен аяқталады. Морфемааралық және сөзаралық шекаралар бұл алгоритмнің жұмысына қатысы жоқ.

Артикуляторлық қозғалыстарды басқару тетігімен шұғылданатын мамандарды адамның ми жұмысының мәліметтері де қызықтыра бастады. Әсіресе, олар ми жұмысының сол жақ және оң жақ жарты шар қабықтарының функционалды асимметриясын зерттеу жұмыстарына қатысты болады.

Зерттеу аясының үлкен бір бөлігі – әртүрлі ми зақымына қатысты сөйлеу тілінің бұзылу проблемалары. Қазіргі кезде ондай артикуляторлық бұзылу сипатына қарай, ауруларға дәл диагноз қоюға мүмкіндік туды.

Сонымен бірге артикуляцияның бала кезінен бастап бұзылу сипатын зерттеу мәселесіне де аса көңіл бөлінуде. Мысалы, логопед-дәрігерлер баланың сөйлеу тілінің бұзылу түрлерін барынша толық қамтитын жүйеге келтірді деуге болады.

Сөйлеу тілінің бұзылу жағдайын мынадай екі типке бөлуге болады:

1) айтылымның фонациялық (сыртқы) берілісінің бұзылу жағдайы, яғни айтылымның сырт түрінің бұзылуы;

2) айтылым берілісінің құрылымды-семантикалық (ішкі) бұзылуы, яғни сөйлеу тілі бұзылуының жүйелік немесе полиморфтық жағы.

Өз тарапынан фонациялық берілістің бұзылуы: дауыс арқылы жасалудың, қарқынды-ритмдік жасалудың, интонация-саздылықтың және дыбысталудың

бұзылуларына байланысты болып келеді. Бұл аталған бұзылу типтері жеке де және аралас сипатта да кездесуі мүмкін.

Енді қолданбалы фонетика мамандарына қажетті деп саналатын логопед-дәрігерлердің қолданысындағы терминдердің түсінігіне қысқаша тоқталайық.

Дыбыс аппаратының патологиялық өзгерісінің әсерінен болатын фонацияның толық қатыспауы немесе оның бұзылуы – *дисфония* немесе *афония* деп аталады; сөйлеу тілінің патологиялық бәсеңдеу қарқыны – *брадилалия* терминімен аталады; патологиялық сөйлеу қарқынының өсуі (тездетуі) – *тахилалия*; сөйлеу мүшесінің (құралының) бұлшық еттері сіңірінің тартылу жағдайына қатысты болатын сөйлеудің қарқын-ритмикалық құрылуы – *тұтығу;* тыңдау мүмкіндігінің қалыпты жағдайында және иннервацияның (нервпен қамтамасыз ету) бұрынғысынша сақталған жағдайындағы дыбысты айту мүмкіндігінің бұзылуы – *дислалия*; сөйлеу аппаратының анатомиялы-физиологиялық дефектеріне байланысты болатын тембр мен дыбысты айту мүмкіндігінің бұзылулары – *ринолалия* деп аталады; сөйлеу аппаратының нервпен қамтамасыз етуінің (иннервация) жеткіліксіз болуынан сөйлеудің айтылу жағының бұзылуын – *дезартария* деген терминмен атайды.

Айтылымның құрылымдық-семантикалық берілу жағының бұзылуы екі түрлі көрініс табады. Оның біріншісі – *алалие*, яғни бастың ми қабығының сөйлеуге қатысты зоналарының органикалық бұзылу салдарынан болатын сөйлеу мүмкіндігінің жоқтығы немесе дұрыс дамымауы. Екіншісі – *афазия*, яғни адамның ми жұмысының тек бір жеріне ғана (локальный) зақым типіне байланысты болатын толық немесе жарым-жартылай сөйлеу мүмкіндігінің бұзылуы.

Міне, осындай балалардың сөйлеу тілінің бұзылуларының негізгі түрлерін үстірт қана атаудан, сөйлеу тілінің сипаттамаларын білудің маңыздылығы айқындалып тұр. Сондықтан логопедия саласы өз пациенттерімен жұмыс істерде фонетикалық мәліметтерді кең пайдаланады.

Артикуляторлық фонетиканың шешілмеген мәселелерінің ішінде дауыс шымылдығы жұмысын зерттеу аясы айтарлықтай орын алады.

Жоғарыда айтылғандай, дауысты, сонант және ұяң дауыссыз дыбыстардың жасалуында дыбыс шымылдықтары (желбезек) дыбыстың қайнар көзі рөлін атқарады. Сонымен бірге, олардың жұмысы ең маңызды интонациялық құрал ретіндегі интонацияның сазды (сарынды) компоненттің де рөлін атқарады. Вокальдық сөйлеу тілі мен балалар сөйлеу тілін зерттеу аяларындағы мамандардың, логопедтердің және т.б. ғылыми ізденістерінің негізінде адам дауысын туындату теориясы дамып келеді. Мұндағы дискуссия туғызатын негізгі мәселелер деп мынадай сауалдарды айтуға болады:

- дыбыс желбезегі (шымылдығы) шеттерінің қабысқан жері арқылы өтетін ауа тогынан пайда болатын тербелістер негізінде және фонация кезінде туындайтын тербеліс процесін пассивтік сипатта деп қарастыруға бола ма?

- немесе керісінше, тербеліс жиілігі дауыс желбезегінің (шымылдығының) икемділік дәрежесіне байланысты, яғни адамның дауыс желбезегі (шымылдығы) ауа тогының әсерінен болатын тербелісі пассивті емес, нерв импульстерінің қадағалау әсерінен активті түрде қысқарып отырады, деп санауға бола ма?

Сөйлеу тілі туындауының акустикалық теориясы тұрғысынан қарастырғанда, артикуляторлық процестер жайлы мәліметтер, сөйлеу тіліндегі хабарлаулардың акустикалық қасиеттерін өңдеу жұмыстары үшін аса маңызды болуда.

4.10 Сөйлеу тіліндегі хабарларды қабылдау мәселесін зерттеу

Адамның есту жүйесі ауа тербелісін нерв импульстеріне ауыстырып, өзгертуіне толық мүмкіндігі бар. Адамның ми қабында орын алатын нерв импульстерінің қасиеттері фонетикалық түсініктеме (интерпретация) алу мен шешім қабылдау жағдайлары үшін пайдаланылады. Адамның есту қабілеті арқылы сөйлеу тілі дыбыстарын өңдеу өте жетілген процесс екендігі белгілі. Сондықтан адамның қабылдау мүшесінің табалдырығындағы сигналдар жүйесін анықтау, тану және оларды теңестіру процестері ақиқат уақыт өлшемі мезгілінде орындалады. Сөйлеу тілін танып-білу жүйесін модельдеумен айналысатын мамандар аталған процестердің қасиеттерін техникалық таным құрылғысын құрастыру кезінде молынан пайдаланады. Осындай жағдайда, адамның есту жүйесіне келіп тірелетін сөйлеу тілі сигналдарын бірнеше өзара айырым табатын сатылар күйінде қарастырып барып, зерттеу қажеттігі туындайды. Ең алдымен, адамның кез келген дыбысты есту қабілеті, соның ішінде сөйлеу тілі дыбыстарын қабылдау қасиеттері жеке қарастырылады. Бұл қасиеттер әрбір есту қабілеті дұрыс қалыптағы жеке адамға және қайсібір нақты дыбысқа тән болып келетін акустикалық сипаттамалардың қатынасы ретінде қарастырылады. Ал мұндай нәтижелер адамның есту аппаратының мәліметтерді бағалау мүмкіндігін білу үшін қажет. Аталған сипаттауларға дыбыстың жиілігі мен оның биіктігі арасындағы қатынас, адамның қатты шыққан дыбысты қабылдауы мен оның жиілігі арасындағы тәуелділікті және т.б. жатқызуға болады. Бұл зерттеулердің барлығы да психоакустика саласына қатысты және олардың маңыздылығы тек адам баласының есту мүмкіндігін қайталайтын техникалық жүйелерді құруда ғана емес, сонымен бірге байланыс арнасы бойынша таратылатын сөйлеу тіліндегі хабарларды таратудың сапасын жетілдіру үшін де өте қажет. Соңғы сөз болған мәселе аса жетілген дыбыс шығарушы аспаптарды (аппараттарды) шығаруда, концерт қойылатын залдардың, оқу аудиторияларының акустикалық деңгейлерін жетілдіру мақсатында да өте маңызды.

Дыбыстарды қабылдаумен айналысатын ғылым саласы **биоакустика** деп аталады. Ал **психоакустика** осы ғылымның бір бөлігі ғана болып саналады. **Биоакустик** мамандардың зерттеулері бойынша, біз тек адамның ғана есту қызметінің ерекшелік қасиеттерін біліп қоймай, басқа да жан-жануарлардың да есту мүмкіндіктерін біле аламыз.

Қазіргі кезде адамның есту қызметінің сөйлеу тілі дыбыстарының сигналдарын қабылдау ерекшеліктері арнайы зерттеу аясына айналып отыр. Кейбір жануарлардың да адамның сөйлеу тілін айыра алады, бірақ соған қарамай, тек адам баласы ғана сөйлеуге қажетті дыбыстық ортаны туғыза алады. Мұндай орта мән-мағынасы бар ақпаратпен өзара алмаса алады. Басқаша айтқанда, тек адам баласы ғана белгілі жүйе құрайтын дыбыстық сигналдарды әрі туындатуға, әрі оны қабылдауға табиғи жағынан мүмкіндігі бар. Мәселен, адам баласының қай тілде сөйлеуіне қарамай, кез келген адамға тән сипаттағы сөйлеу тілі сигналдарын өңдей алатын әмбебаптық әдіс-тәсілдері бар. Адам баласының есту қабілетінің әмбебаптық қасиетіне мынадай жайттарды жатқызуға болады:

- дауысты мен дауыссыз дыбыстарды ажырату мүмкіндігі;
- сөйлеу тілі хабарының элементтері ішінен қажетті акустикалық ақпаратты бөліп алу мүмкіндігі;
- дыбыстың қоршауындағы элементтердің акустикалық қасиеттеріне қарай, түсіп қалған (жоқ) элементті қайтадан қалпына келтіру мүмкіндігі;

- дыбыстардың кездесу ықтималдығын ескере білу және т.б. адамның ойлау және сөйлеу қызметінің даму процесінде қалыптасып қалған көптеген ерекшеліктерді ескере білу мүмкіндігі.

Мұндай сөйлеу тілі сигналдарын қабылдай алу мүмкіндіктерінің ерекшеліктерін білу – зерттеудің психофизиологиялық мәліметтерін айқындау үшін аса маңызды. Себебі психофизиологиялық деректер адамның жоғарғы нерв қызметінің күрделі механизмдеріне байланысты мәселелерді қозғайды.

Қабылдаудың қасиеттері ішінен тілдің нақты дыбыстық жүйеге қатысты және жеке индивид үшін ана тілі болатын тілдік қасиеттердің тұжырымдаулары айрықша назар аударуға тұратындай. Адамның есту мүшесі арқылы сөйлеу тілі дыбыстарын өңдеу тәсілдері дыбыс элементтерінің санына байланысты болады. Мысалы, орыс тіліндегі сөйлеу тілін қабылдау үшін кейбір акустикалық параметрлер бойынша кемінде әртүрлі алты дауысты және 36 дауыссыз дыбыстарды теңестіру (индентифицировать) керек, ал француз тілін қабылдау үшін ондай мәлімет кемінде он бес дауысты және жиырма дауыссыз дыбыстарға тең. Қазақ сөйлеу тілінің төл дауысты дыбыстарының жүйесі 9 дауысты дыбыстан, ал төл дауыссыз дыбыстарының жүйесі 17 дыбыстан құралады (егер ***К*** мен ***Г*** дауыссыздарын қоссақ, онда – 19) [Жүнісбек, 2005, 5-б.]. Дыбыс элементтері санының мұндай өзара айырым белгілері дауысты және дауыссыз дыбыстардың артикуляторлы-акустикалық қасиеттерінде де қатты айырым табуы керек. Қолданбалы фонетика үшін бұл мәселе аса маңызды болып саналады. Мысалы, қайсыбір дыбыстық жүйеден екінші жүйеге өтетін жағдайда дыбыстарды теңестіру процедурасы қалай өзгереді деген сауал туындайды. Ең бірінші, бұл мәселе шет тілдерін оқыту бағытында маңызы зор. Егер оқушылар шет тілінің дыбыс айырмашылығын дұрыс қабылдап жатқанына сеніміміз мол болса ғана, сол тілдің артикуляциясын дұрыс үйрету мүмкіндігі туады. Қабылдаудың тілдік ерекшелігі автоматтанған танымдық жүйесін және сөйлеу тілін синтездеудің техникалық жүйелерін құру кезінде де ескеріледі.

Қабылдауды зерттеудің негізгі әдістері. Қабылдау ерекшеліктерін зерттеуге қатысты тәжірибелерде сынаққа қатысушыға дыбыстық материалдардың қандай қасиеттерін және қандай есеп шартын ұсыну жолдарын айыра білу аса маңызды. Практикалық тұрғыдан келгенде зерттеушіні қызықтыратын негізгі мәселелер мыналар: дыбыстық бірліктердің қандай акустикалық сипаттамалары және нақты қандай мәселенің шешімін табу үшін олар маңызды. Осыған байланысты "стимул" ретінде қандай тілдік бірліктер дыбыстарын алған жөн екендігін алдын ала шешіп алу қажет. Мысалы, "стимул" ретінде табиғи тілдегі сөзді алатын болсақ, оны біз қалай бұзып айтсақ та, көп жағдайда, оны тану қиынға түспейді. Себебі, ол ең алдымен адамның қай жағдайда да дыбыстар ағымы ішінен сөзді бөліп алуға ниеттенуінен де болуы немесе әрбір сөздің оны танып-білуді жеңілдететін қосымша сипаттамаларға ие болуынан деп те түсіну керек.

Тәжірибе кезінде ұсынылатын **стимулдар** (тірек ақпарат) әртүрлі созылыңқылықтағы табиғи сөйлеу тілінің сигналдары немесе синтезделінген сөйлеу тіліне ұқсас болып келетін сигналдар. Соңғы аталған **стимулдардың** сипаттамалары тәжірибешінің қалауы бойынша алынады және оның қадағалауымен өзгеріске ұшырауына мүмкіндігі бар. **Бірінші стимул** ретінде, яғни табиғи сөйлеу тілінің сигналдарына материал болатын магниттік таспаға жазылып алынған сөйлеу тілі. Ал бұл тілдік материалдан тәжірибешіні қызықтыратын дыбыстық сигналдар бөліп алынып, қайтадан таза магниттік таспаға көшіріледі. Осылайша бөлініп алынған сегменттер сынаққа қатысушыларға ұсынылады. Қазіргі кезде осы айтылғандардың бәрі арнайы бағдарламалар арқылы компьютер көмегімен жүргізіледі. Ондай

бағдарламалар арқылы сөйлеу тілі сигналдарының цифрлық пішіні бойынша қажетті дыбыстарды сегменттеу процесі орындалып, зерттеу мәселесіне қатысты тәжірибе іске асады. Компьютерлік техника алдын ала берілген параметрлері негізінде синтезделінген стимулдарды жасап, эксперимент барысында оларды (табиғи және жасанды) өзгеріске ұшыратып отыруына мүмкіндігі бар.

Енді тәжірибе кезінде сынаққа қатысушы қандай мәселелерді шешетіндігіне қысқаша тоқталайық.

Тірек ақпаратты (стимулды) тану. Бұл тәжірибелерде сынаққа қатысушы ұсынылған тірек ақпаратты қабылдай сала, "нені" естіп тұрғанын анықтауы қажет. Әдетте мұндай жағдайда сынға түсуші адам өз жауабын арнайы анкеталарға жазып отырады. Мұндағы негізгі қиындық – сәйкес дыбысты дұрыс таңдай білу.

Қос-қостап салыстыру. Бұл әдіс бойынша сынаққа қатысушылардың ұсынылған тірек ақпараттарды (стимулдарды) олардың белгілі акустикалық сипаттамалары бойынша ұқсас немесе ұқсас еместігін айыра білу қабілеті анықталады. Салыстыруға ұсынылатын **стимулдар қос-қостап** беріледі. Сынаққа түсуші тек екі стимулдың арасында ұқсастық **"бар"** не **"жоқ"** деп жауап беруі қажет. Ал дыбыстың таңбасын анықтаудың қажеттігі болмағандықтан бұл тәжірибе қиындық туғызбайды.

Іздеудің белсенді әдісі. Бұл әдіс қос-қостап салыстыру әдісінің бір түрі. Әдетте мұндай тәжірибе синтезделінген дыбыстармен жүргізіледі. Ұсынылатын синтезделінген дыбыс тұрақты сипаттамаға ие болады, ал сынаққа қатысушыға, осыған ең көп ұқсас не ең көп айырым табатын стимулды (тірек ақпаратты) таңдап алуына тапсырма беріледі. Жауап беруші синтезатор арқылы стимулдың сипаттамаларын өзгертіп отырып, ұқсас не ұқсамайтын стимулды есту қабілетінің негізінде іздеп табуы керек. Бірақ ол өзі қандай өлшемдерді (сипаттамаларды) өзгертіп отырғанын білмейді, тек есту мүмкіндігі арқылы есептің жауабын табуы керек.

ABX* – салыстыру.** Бұл тәжірибеге үш стимул қатынасады. **(*A, B* және *X*). *A және ***B*** стимулдарының сипаттамалары белгілі. Осы екі белгілі стимулмен үшінші белгісіз стимулды салыстыра отырып, оған тиісті баға берілуі қажет, яғни сынаққа түсуші **үшінші *X* стимулын** анықтауы қажет. Қос-қостап салыстыру әдісіндегідей бұл әдісте сынаушы стимулдың таңбасын іздеуден бастайды. Жоғарыда аталған әдіске қарағанда, бұл әдістің шешімге әсерін тигізетін акустикалық параметрлердің шамаларын бағалауға мүмкіндігі мол.

Еліктеу (имитация) әдісі. Бұл әдіс бойынша сынаушыға бірнеше стимулдар ұсынылады. Ол қабылдап алған стимулдарын қайталап айтып шығуы керек. Қайталанып айтылып жатқан уақытта айтылған дыбыстар магнитты таспаға жазылып, оларға тиісті акустикалық талдаулар жүргізіледі. Алынған мәліметтер бойынша сынаушының еліктеп қайталаған дыбыстың қай белгілері сынаққа түсушіге тән екендігі анықталады. Еліктеу арқылы зерттеу тек сөйлеу тілі сигналдарын есту процесін өңдеуді ұғыну үшін ғана қажет емес, сонымен бірге айтылымды туындату кезіндегі сөйлеу тілі қызметіне бағдарлама жасаудың мәнін терең ұғыну үшін де аса қажет.

Психофизиологиялық шкалалау. Шкалалау әдісі **классикалық психоакустика** шеңберінде пайда болып, сөйлеу тілін техникалық құралдармен зерттеудің даму кезеңінде жаңа қарқынға ие болды. Шкалалау әдісі сигналдардың акустикалық қасиеттерімен және оларды есту жүйесі арқылы бағалау арасындағы тәуелділікті айқындау үшін қолданылады. Қазіргі кездегі шкалалау теориясы екі түрлі қатынастарды негіз етіп қарастырады. Олар – **жақындық** қатынасы және **артықтық**

қатынасы деп аталады. **Жақындық** түсінігі элементтер арасындағы арақашықтық жайлы субъективті ұғымға ауысуға мүмкіндік жасайды. Ал **артықтық** қатынасы сынаушының, көптеген жағдайда, кейбір субъективті анық белгілер бойынша екі сигналдың арақатынасын анықтай алу мүмкіндігіне негізделеді.

Сөйлеу тілінің дыбыстық тізбегін қабылдау тетігін зерттеу осындай жалпыпсихологиялық алғышарттарға сүйенеді.

Шкалалау теориясы шеңберінде орындалған нақты жұмыстарда ұсынылған **сөйлеу тілі стимулдарына** қатысты сынаушы мынадай мәселелерді шешеді:

1) стимулдар арасындағы **субъективті арақашықтықты** анықтау. Мұндай жағдайда шкала шамасы эксперимент жүргізушінің өз қалауымен қойылады;

2) **субъективті жақындықтың** деңгейі стимулдарды есте сақтаудың серияларымен анықталады. Стимулдарды есте сақтау нәтижелері негізінде бір стимулды екінші стимулмен алмастырудың ықтималдығы анықталады.

3) **семантикалық қарама-қарсы жұптар**. Қарама-қарсы мағынадағы жұп сөздермен беріліс жағдайында, әрбір стимул қайсыбір белгілер жиынына қатысты бағаланады. Сынаушы әрбір стимулдың орнын шеткі белгілердің мәніне қарай анықтайды.

Барлық тәжірибе нәтижелерін шкалалау дегеніміз, ол күрделі статистикалық аппарат арқылы алынған мәліметтерді өңдеу және орындалатын амалдардың математикалық модельдеріне сүйену.

Сөйлеу тілі әрекетінің фонетикалық аспектілері. Сөйлеу тілінің дыбыстық жағы лингвистикалық тұрғыдан тілдік мағыналарды таратудың тұрпаты ғана деп сипатталады, сондықтан өзіне тән мән-мағына жүгін көтере алмайды деп түсіндіріледі. Ал егер адам баласы өзінің мүмкіндігіне қарай, бір мезгілде әр аяда әрекет ете алатын күрделі жүйе екендігін ескеретін болсақ, мәселе басқаша түсінік алады. Сөйлеу тілі қызметінің кейбір бағыттарында сөйлеу тілінің тек дыбыстық формасы ғана ең басты рөл атқарды. Мәселен, ең алдымен, олар ырғақты сөйлеу тілі – өлеңдер, аңыздар, ертегілер, батырлар жырлары. Аталғандардың мазмұнын дұрыс түсіну үшін кейбір дыбыстардың орынды қайталануы соншалықты маңызды болып саналады, тіпті мағыналық бірліктің алғашқы бейнесі өзгеріске ұшырап, жаңадан фонетикалық бейне пайда болады. Ондай фонетикалық бейне – сөздердің кейбір бөліктерін қайталау арқылы, сөз алдынан не артынан неше түрлі қосымшаларды, қыстырмаларды қосып, дауысты не дауыссыз дыбыстарды созып айтып және т.б. жолдармен жасалады.

Әсер етудің арнайы түрі дегеніміз, ол сөйлеу тілін басқаша – әдеттен тыс және осы тілге жат түрде құру деп түсінуге де болады. Аталған жайт, **"фонетикалық мағына"** деген құбылысқа негізделеді. Оның мәні мынада. Адам баласы дыбыстарды материалды ғаламның дербес құбылысы деп қарастыруға және дыбыстардың әрбіреуіне белгілі баға беруге икемді болуынан. Мысалы, барлық адамдар үшін мынадай *[а]* және *[і]* қос дыбыстардың біріншісін қайсыбір үлкендік мағынаға тән, ал екінші *[і]* дыбысын, керісінше, қайсыбір кішкенелік мағынаға ие деп ұғынудан. Сол сияқты, сонант дыбыстарды қайсыбір ***"сүйкімді, жағымды, тәтті, ұнамды, жүзі жылы, көз тартатын, сымбатты"*** ассоциациалармен, ал шуыл дауыссыз дыбыстардың **"жағымсыз"** мағынада ассоциаланатыны белгілі.

Дағдылы мәтіндерде дыбыстардың кездесу ықтималдығы статистикалық заңдылықтармен анықталады. Бұл заңдылық нақты тілдің дыбыстық құрамы мен құрылысын сипаттайды. Автор өз шығармасы арқылы оқырманға не сөйлеуші, өз айтылымы бойынша тыңдаушыға әдейі әсер қалдыру мақсатымен дыбыстарды

орналастыруда әдеттегіден айырықша құрастырады. Егер қайсыбір мәтінде **"жағымды"** дыбыстар көптеп кездессе, онда ол мәтін мазмұны оң эмоциялы бояумен қабылданады, ал егерде онда **"жағымсыз"** дыбыстар басымшылық етсе, қабылдау сипаты да кілт өзгереді, яғни жағымсыз эмоциямен қабылданады.

Мәтін иесі сөйлеу тілінің тыңдаушыға әсерінің сөздің фонетикалық безендірілуіне байланыстылығын ертеден-ақ біліп, оны өте орынды қолданғандары мәлім. Мысалы, шешендік өнері теориясы – риторикаға қатысты ертеректегі арнайы оқу құралдарында, сөйлеу тілінде қалаған эффектіге жету үшін дыбысталу тілімен мұқият айналысу қажеттігі жайлы ескертулерді кездестіруге болады.

Сахналық сөйлеу тілі көрермендерге фонетикалық қасиеттер арқылы нағыз әсер қалдыратын арнайы алаң деуге болады. Сахналық сөйлеу тілінің өзіне тән көптеген ерекшеліктері тек сахнада ғана емес, сонымен бірге барлық қоғам мүшелеріне де тигізетін әсері ерекше. Себебі, сахналық тіл идеалды айтылу нормасын іске асырып, нақтылы түрде көрсетеді және сол тілдің әдеби сөйлеу тілінің үлгісі ретінде қабылдануы керек. Бірақ сахналық сөйлеу тілі қызметінің шындығына көңіл аударатын болсақ, идеалды және ақиқаттық сипаттары арасында көптеген қарама-қарсылықтар жоқ емес.

Ең алдымен, сахналық сөйлеу тілі – ол, тілдік жағдайды бейнелеуге арналған еліктеу тілі. Сондықтан тілдің табиғи жағының ұғымы тек жарым-жартылай деуге болады. Сахна тілінің айрықша мәнерлілігі, негізінен, сахна шеберінің қалаған әсер деңгейіне жету ниетінен туындайды және ол маманданған актерлік дағдымен де байланысты болуы мүмкін.

Актерлік сөйлеу тілінің техникалық жағы орындаушының арнайы фонациялық демалысының, дыбыстың жылдамдық (тездік) сипатының және дұрыс қойылған дикциясының ерекшеліктеріне тікелей қатысы бар. Театр өз тұрғысынан, орфоэпиялық норманы сақтаушы бола тұра, ол сонымен бірге, персонаждың сөйлеу тілінің бейнесін сақтау үшін, диалектінің фонетикалық ерекшеліктеріне де жүгінеді.

Радио дикторларының сөйлеу тілінің ерекшеліктерін теледидар дикторларының тілімен салыстыра зерттеумен бірге, саяси жаңалықтарды, көңіл көтеру бағдарламасын, спорт жайындағы репортаждарды және т.б. әртүрлі жанрлық сөйлеу тілінің дыбыстық ерекшеліктерін зерттеу қолданбалы фонетиканың маңызды мәселелерінің бірі болып отыр.

Бұқаралық ақпарат құралдарының қоғам өмірінде үлкен рөл атқару жағдаяты да сөйлеу тілін зерттеушілерге аса маңызды мәселелерді шешуді талап етуде. Сөйлеу тілі арқылы қоғамға әсер етушілерге айтылымның фонетикалық қасиеттері мен сипаттамаларының маңыздылығын ескеру әрдайым қажет. Бұл тек сөйлеуші тілінің мәдениетін тәрбиелеу үшін ғана емес, сонымен бірге әдеби тілдің дамуы мен оның дыбыстық құралдарының өмір сүретін фонетикалық ортаны "ластаудан" сақтаудың бірден-бір дұрыс жолы деуге болады.

Бесінші тарау
ТІЛ БІЛІМІНДЕГІ СТАТИСТИКАЛЫҚ МОДЕЛЬДЕУ ӘДІСТЕРІ

5.1 Статистикалық әдістің танымдық принциптері

Қазіргі кезде филологиялық ғылымдар саласындағы статистикалық әдіс-тәсілдердің және оның қандай нақты есептерді шешудегі орны мен рөлін анықтау мәселесі алдыңғы қатарға қойылуда. Статистикалық әдістің басқа да әдістерге қарағанда қандай мәселені шеше алатындығы немесе шеше алмайтындығы жайлы жайттың басын ашу қажеттігі әлде де өз құндылығын жойған жоқ.

Статистикалық әдіс дегеніміз – ол, ғылыми теориялық және практикалық қорытындылар жасау үшін статистикалық деректерге жүргізілетін топтау, жүйелеу, өңдеу және түсініктеме беру мақсатымен орындалатын неше түрлі әрекеттер. Туындаған статистикалық жағдайға қарай осы аталғандарды іске асыру актуальды мәселеге айналып отыр. Бұл жердегі статистикалық берілістер немесе деректер деп отырғанымыз – зерттеу нысанының жиынтығы жайлы сандық мәліметтер. Зерттеу нысаны – үнемі сандық және сапалық қалыптары өзгеріске ұшырап отыратын және ортақ қасиетке ие болатын зат не құбылыс (нысан). Әдіс-тәсілдер жиынының математикалық мазмұны мен статистикалық әдістің принциптік жағы қолданбалы математика саласының математикалық статистика пәніне жатады. Сонымен бірге статистикалық әдіс статистикалық зерттеудің логика-танымдық проблемалық мәселесіне қатысты математикадан тыс жақтарын да қамтиды. Олардың бірқатары әмбебаптық сипатқа ие, ал басқалары нысанның сипаты мен әрбір ғылыми пәннің ерекшелігіне қарай анықталады.

Статистикалық әдістің филологиялық мәселелерді шешуге бейімделген жақтары, негізінен, екі бағытта іске асады. Оның біріншісі, филологиялық, әсіресе лингвистикалық терминдер тұрғысынан ықтималды-статистикалық категорияларды қайта қарастырып, оған жаңа түсінік беруге талпыну болып табылады. Яғни тілдік бірліктерге қатысты ықтималдық, жиілік, кездейсоқ шама, жиынтық, таңдама бөлік, корреляция, топтау және т.б. Осы айтылғандар сияқты ықтималдықтар теориясы мен математикалық статистика терминдерінде тілді зерттеу тұрғысынан қайта қарастыру қажеттігі. Екінші бағытқа қатысты тән қасиет – бірінші бағытқа қарағанда ой әрекетінің керісінше сипаты. Мәселен, орыс тіліндегі әдебиеттерде қолданыс табатын мынадай: “язык” және “речь”, “норма” және “узус”, “текст” және “стиль”, “план выражения” және “план содержания”, “поверхностный уровень” және “глубинный уровень”, “прагматика” және “синтагматика” және т.б. негізгі лингвистикалық категорияларды ықтималды-статистикалық тұрғыда қайта қарастырып, жаңаша ой түйініне келу.

Филология-статистикалық және статистика-филологиялық деп аталатын екі түрлі тенденция аралығына анық шекара сызу қиын. Әйтсе де, ол принциптік мағынаға ие емес, себебі бұл жерде тіл білімінің категориялы-ұғымдық аппарат пен статистиканың бір-бірімен табиғи түйісу фактісінің өзі маңыздылау деп саналады. “Ықтималдық” терминімен аталатын ықтималды-статистикалық категория

тіл білімі саласында "лингвистикалық ықтималдық" терминінің әртүрлі варианттары арқылы көрініс табады. Олар: "тілдік ықтималдық", "сөйлеу тілінің ықтималдылығы", "парадигматикалық ықтималдылық", "синтагматикалық ықтималдық", "нормалаушы ықтималдық" және т.б.

Статистикалық әдіс – ақиқат шындықты танудың әмбебап әдісі. Танымдық сипат – субъект пен зерттеу нысаны (объект), зерттеу нысаны мен әдіс, нысан мен ол туралы білім қоры араларындағы қарым-қатынасты анықтау арқылы іске асады. Тану процесі зерттеушінің объективтілік пен алалыққа жатпайтын принциптерін ұстаным етумен сипатталады.

Белгілі бір нысанды зерттеу объектісі орнына алу, ең алдымен, практикалық мұқтаждықтан туындайтын және қолданбалы мақсатқа сай келетін прагматикалық тұрғыда жүргізіледі.

Зерттеушінің аталған нысанмен өзара іс-әрекеті әрі нақтылықпен, әрі жуықтықпен сипатталады. Бұл қасиеттер бір-біріне қарсы келмейді, себебі статистикалық әдістің негізінде ықтималдықтар теориясы жатыр. Сондықтан барлық пайымдаулар математикалық тұрғыда қатаң бола тұра, ықтималдық, жуық сипатта түсінік табады. Статистикалық әдістің белгілі тәртіппен орындалу сипаты зерттеушінің объективтілік ұстанымын нығайта түседі. Әдістің жуықтық сипаты танымдық мақсаттың прагматикалық бағытымен тығыз байланысты. Оның себебі қолданбалы мақсаттағы тілге қатысты нақты есептерді шешуде оны жан-жақты сипаттаудың қажеттілігі болмай, көбінде, жуықтап сипаттаумен де қанағаттануға болады.

Зерттеу нысаны қатаң детерминделінген жүйе емес, керісінше, жұмсақ, диффузды, ықтималды жүйе. Сондықтан ол белгілі уақыт аралығы мен кеңістіктік және көптеген белгілер топтамасын қамтитын бірліктер жиыны. Зерттеушіні осындай орта ішінен ақиқат белгілерден гөрі, сандық табиғатқа ие симптомдық белгісі бар нышандар көбірек қызықтырады.

Қызығушылық ұстаным және әдіс тұрғысынан келгенде ізденушінің зерттеу нысанымен өзара әрекетінен белгілі дәрежедегі статистикалық білім туындайды. Ондай білімге тән қасиеттер мыналар [Мартыненко, 1996, 201-223-бб.]:

а) зерттеушінің ұстанымы мен қызығушылығына қатысты – объективтілік, прагматикалық;

ә) зерттеу нысанына қатысты – жүйелілік, стохастылық, симптомдық, көпөлшемдік, нақтылық, сандық;

б) зерттеу әдісіне қатысты – қатаңдық, жуықтық, теориялық жайтты іс жүзінде тексеру мүмкіндігі (верификация).

Осы соңғы аталғанға қатысты мынадай жайтқа да көңіл бөлінгені жөн болады. Зерттеу әдісі тек зерттеу нысанымен ғана өзара әрекеттесіп қоймайды, сонымен бірге ол қол жеткен білім дәрежесімен де әрекеттеседі. Осыдан барып, білімнің верификациялану қасиеті туындайды, яғни ақиқаттықты тәжірибе жүзінде дәлелдеу немесе теориялық ақиқаттықты тексеру мүмкіндігі туындайды.

Статистикалық танымдық принципі көптеген жағдайда ортодоксалды (бұлжытпай бір бағытты ұстану) дискриптивтік лингвистикамен үндесіп жатады. Егер дискриптивті лингвистиканың негізгі әрекет жасау ортасы дистрибуция (тілдік бірліктердің тіркесімдігі) екенін, ал сипаттауға негізделген статистиканың әрекет ортасы – ықтималдық үлестірілу екенін ескерсек, онда олардың бір-біріне ұқсастық дәрежесі арта түседі. Мысалы, "дистрибутивті-статистикалық әдіс" деген қос сөздің компоненттері бірін-бірі қайталап тұр деуге болады. Себебі, кез келген статистикалық әдіс диструбутивті болады. Ал терминдердің басқаша қолданысы – "статистика-комбинаторлық әдіс" өте орынды, себебі кез келген ықтималды-статистикалық модель комбинаторикаға негізделеді.

5.2 Лингвистикалық мәліметтерді өңдеудің негізгі әдісі

Дәстүрлі статистикада екі түрлі әдіс топтамалары айырым табады: сипаттау әдісі және бағалау әдісі.

Алғашқысының міндеті – тілдік құбылыстарға қатысты бастапқы мәліметтерді ықшамды және көрнекі пішінде (арнайы кесте, сызба арқылы) көрініс тапқызу. Сосын барып, әртүрлі статистикалық тәсілдер негізінде жинақталған сандық мәліметтерді қортындылау. Оларға жататындар: орталық түйінді ойдың өлшемі, өзара тәуелділіктің өлшемі және т.б.

Сипаттау әдісі бірдей дәрежеде тәжірибелік деректермен әрекет ететін және ықтималды-статистикалық модельді таңдау тұрғысынан қарастырғанда, модельдеу әдісіне іргелес (жақын) орын алады.

Статистика саласында ұзақ уақыт бойы нормальды (гаусстік) үлестірілу заңы басымдық етті, содан кейін барып басқа да үлестірілу заңдары қолданыла бастады. Мысалы, "нормаль үлестірілуіне" жақын жататын (пуассондық, биномиальдық, логарифмды-нормальдық) заңдылықтар зерттеу "сахнасынан" орын ала бастады. Қазіргі кезде "нормальды үлестірілу" заңының зерттеу аясынан алатын орны барынша кішірейді деуге болады. Себебі, әлеуметтік процестерді модельдеуде "гаусстік емес" статистикалық үлестірілу концепциясы басымшылыққа ие бола бастады [Яблонский, 1975; Хайтун, 1973; Мартыненко, 1987). Осы жөнінен Г.Херданның батыл пікірлерін айта кетуге болады. Оның тұжырымдауынша, фонологиялық деңгейдегі статистика мен сөздік деңгейіндегі статистика мүлде бөлек тұрған математикалық пәндерге жатады. Фонемдік статистика математикалық статистика мен ықтималдықтар теориясының классикалық аппаратына сүйенеді, ал сөздік статистикасы – статистикалық физика саласынан алынған комбинаторлық модельдерге жүгінеді (Herdan, 1964).

Бағалау әдістері де екі топқа бөлінеді: үлестірілудің белгісіз өлшемдерін бағалау әдісі және статистикалық болжамдарды тексеру әдістері. Әдістердің бірінші тобы зерттеуге қатысатын шамалардың ықтималдыққа қатысы барын ғана қарастырады. Ол, "үлкен сандардың заңдылығы" деп аталатын заңдылыққа және классикалық шекті теоремаларға (Чебышев теоремасы, Ляпуновтың орталық сипаттағы шектеулі теоремасы) сүйенеді. Бағалаудың классикалық теорияларының келесі орнында, үлестірілудің параметрлері үшін сенімділік интервалын құрайтын нормальды үлестірілу заңдылығы тұр. Параметрлерді бағалау тәсілінің қайсыбірін таңдап, қолдану, көбінде, зерттеу жиынтығының табиғатына негізделеді.

Егер бас жиын көлемі тәжірибеде шексіз көп шамаға тең болса, ондай мәтінді зерттеу әрекеті "өкілдік" сипаттағы таңдама бөліктер арқылы жүргізіледі. Осы таңдамалар негізінде ықтималдықтар теориясында қарастыратын шекті шамаға ұмтылатын статистикалық "бағалау" дәрежесі есептеліп шығарылады. Осындай жағдайда, үлестірілу заңдылығының параметрлері тек бас жиын бойынша ғана "бағаланып" қоймайды, сонымен бірге, кез келген басқа да жиындар бойынша қарастырылады. Бас жиынды "бағалаудың" айрықша жағдайы деп экспресс-әдісті қолдануды атайды. Бұл әдіс ең кіші көлемді таңдама бөліктер арқылы жүргізіледі.

Бас жиын шекті көлемдегі бөлшектелген жиындар құрамы арқылы көрініс табатыны белгілі. Әрине, егер **бас жиын** көлемі айтарлықтай көп болмаса, оны таңдама бөліктер арқылы зерттеудің қажеттігі болмайды.

Соңғы кезде статистикалық болжамдарды сынау әдістеріне қатысты параметрлікке жатпайтын орнықты әдістер аса жиі қолданыс таба бастады. Бұл әдістер зерттеу жиынтығының табиғаты туралы және олардың қандай үлестірілу

заңдылығына бағыну мүмкіндіктері жайлы жорамалдың мардымсыз болу жағдайында қолданылады. Бұған мысал ретінде **дисперсиялық талдауды** атауға болады. Лингвистикалық мәліметтерді өңдеу әдістері жайлы үлестірілудің әртүрлі типтерін тіл білімі саласында қолдану нәтижелеріне талдау жүргізу арқылы ғана толығырақ білуге болады.

5.3 Лингвистикалық үлестірілулердің типологиясы

Статистикалық үлестірілу қатарлары деп сандық мәнге ие болатын реттелген жиынды атайды. Ол белгілі бір нысанаға қатысты өлшемдердің реттелген жиыны болуы мүмкін. Әдетте ондай өлшемдер лингвистикалық бірліктердің статистикалық "салмағына" (мысалы, жиілігіне) тең болады.

Статистикалық үлестірілулердің теориялық маңыздылығы мынада. Табиғи тілдер мен мәтіндер ықтималдық жүйелерге тән бірнеше принциптік қасиеттерге ие болады. Олар, бір типті оқиғаларға тән болып саналатын: реттілік пен ретсіздіктің, тұрақсыздық пен тұрақтылықтың, тәуелділік пен дербестіктің тұтастық сипаттары. Ал нақты зерттеулердегі үлестірулер туралы айтатын болсақ, олар эксперименттік мәліметтердің шағын түрде көрініс табуында, талдау жүргізуде, қорытындылау мен түсініктеме беруде аса ыңғайлы болуымен бірге, олар міндетті түрдегі зерттеу құралына айналып отыр.

Әрі қарайғы баяндауымызда лингвистикалық үлестірілулердің типологиясы жайлы сөз болады.

"Тип құрастыру" белгілерінің орын алуы, біріншіден, тілдік ақиқаттық жағдайдың жүйелілік сипатымен түсіндіріледі. Ал, екіншіден, оның формалды-математикалық сипаты басшылыққа алынады.

5.3.1 Ықтималдықтар теориясы және лингвистикалық ақиқаттық

Мәліметтерді топтаудың көрінісін білдіретін эмпирикалық қатар мен ақиқаттық үлестірілу заңдылығын сипаттайтын теориялық үлестірілулер жайлы тек қарама-қарсылық аясында (осы тақырыпшаның аталуы тәріздес) ғана сөз етуге болады.

Теориялық үлестірілулер заңдылықтары сан жағынан барынша көп болғанымен, эмпирикалық қатарларды тексеруде олардың ішінен ұтымды қолданыс тапқан теориялық түрлері ғана таңдалып алынып жүр. Бұлардың көпшілігі статистика пәнінде айрықша орын алады. Оның себебі, теориялық үлестірілулердің математикалық қасиеттері қолдануда аса қолайлы болып келеді. Теориялық үлестірілулердің ең бір маңызды қасиеті, олардың көпшілігінің бір-бірінен туындайтын жекелік сипатта болуы.

Нақты үлестірілу тобын таңдау шартының алғашқысына эмпирикалық үлестірілуді аппроксимациялау (жуық шамасын табу) қасиетінің болуы, ал екіншісіне – белгісіз параметрлерді бағалау мен статистикалық болжамды сынау мүмкіндігінің болуын жатқызады. Мысалы, ондай шарттарға сай келетіндер: Стьюдент, Пирсон, Фишер және т.б. үлестірілулер.

5.3.2 Гаусстік және гаусстік емес үлестірілулер

Теоретика-ықтималдық тұрғыда қарастырғанда барлық ақиқат және мүмкін болатын үлестірілулер түрлері екі топқа бөлінеді: гаусс тобына қатысты үлестірілулер және оған қатысы жоқ, басқа үлестірілулер, басқаша айтсақ, "гаусстікі емес" деп аталатын үлестірілулер тобы. Бірінші топқа көрсеткіш (экспоненциальды) функцияның қайсыбір түрімен сипатталатын теориялық үлестірілулер тобы жатады. Бұл топтың эталоны ретінде нормальды үлестірілу тобы алынады. Бұған барлық гаусстік топқа қатысты үлестірілулер: биномиальды, Пуассон, Стьюдент және т.б. көптеген үлестірілулер. Екінші топтың эталоны ретінде қабырғалары бір-біріне тең болмайтын гиперболалар теңдеуі алынады. Бұл теңдеу – гипербола тәріздес үлестірілулер типіндегі ең бір идеалды жағдайдың үлгілі асимтотикалық (шексіз жақындай түсетін) түрі деп саналады.

Ықтималдықтар үлестірілулерін реттеп отыру үшін теоретика-ықтималдық комбинаториялық сызбаның қатысы қажет. Міне, осы тұрғыда алғанда, яғни қандай сызбаның қатысы болуына қарай, үлестірілулер бір-бірінен ажыратылады. Кейбір үлестірілулер, мысалы, нормальды, бинормальды, пуассондікі, паскалдікі қатаң теоретика-ықтималдық негіздемеге ие, ал басқаларының (Вейбулл үлестірілуі) ондай негіздемелері жоқ немесе оған жартылай ғана ие болады (мысалы, Ципф үлестірілуі).

Кейбір үлестірілулер топтамасы өзімен функционалды қатынаста тұратын эталон ретіндегі (мысалы, нормальды) үлестірілуінің өзгеріске ұшыраған түрі арқылы құрылады. Мысалы, оған жататын Кэмптейн үлестірілулер топтамасы.

Сол сияқты, үлестірілулер топтамасы бір түрдегі немесе тіпті бір-бірінен қатты айырым табатын бірнеше үлестірілулерден "қысылған" композициясынан тұруы да мүмкін. Мысалы, біркелкі "қысымға" ие болатын бірнеше үлестірілулер композициясы нормальды үлестірілу заңдылығына жақын жатады.

Теориялық үлестірілулер анық түрде **модельденуші** және **критериальды** үлестірілулер болып екіге бөлінеді. Бұған ұмтылатын барлық гипербола тәріздес үлестірілулердің түр-түрі (Ципфа – Парето, Эсту – Лотки, мондельброт және т.б.). Гаусс типіне жатпайтын үлестірілулердің айрықша бір белгісі ретінде кез келген дәрежедегі "моменттердің" шексіздік сипатын айтуға болады. Оның жеке түріне **дисперсияны** да жатқызуға болады.

Гаусстікіне басқаша үлестірілулерді зерттеудің тарихы беріде ғана қолға алынды. Ондай үлестірілулердің негізгі қасиеттерін алғаш рет ғалым П.Леви зерттеген болатын. Ғалымның тұжырымдауы бойынша, теоретика-ықтималдық тұрғыда қарастырғанда гаусстікіне жатпайтын үлестірілулер тұрақты класс үлестірілулеріне жатады.

5.3.3 Типтік және типтік емес эмпирикалық қатарлар

Теориялық үлестірілулерді гаусстікі және "гаусстікі емес" деп қарастыру, дәстүрлі статистикадағы эмпирикалық қатарларды типтік және типтік емес деп ажыратуға сәйкес келеді.

Мазмұндық жағынан алғанда статистикадағы бұлайша екіге бөліп қарастырудың барлық құбылыстарды **типтік** және **жекелік** деп баяндауға қатысы бар. **Типтік** құбылыстарға тән қасиет, оларда тұрақты себептердің басымшылық етуі, ал екіншісінде, яғни **жекелікте** – кездейсоқ себептер шешуі рөл атқарады.

Математикалық көзқараспен қарастырғанда, **типтік қатарлар** – тұрақты мен кездейсоқ себептердің өзара әсер ету мүмкіндіктерінің қатал түрде немесе, ең болмағанда, жуық шамада, үлкен сандар заңдылығы бойынша жинақталып қорытылған жаппай процестердің моделіне сәйкес келуі. Ал барлық басқа қатарлар типтік емес қатарлар тобына қатысты деп қарастырылады.

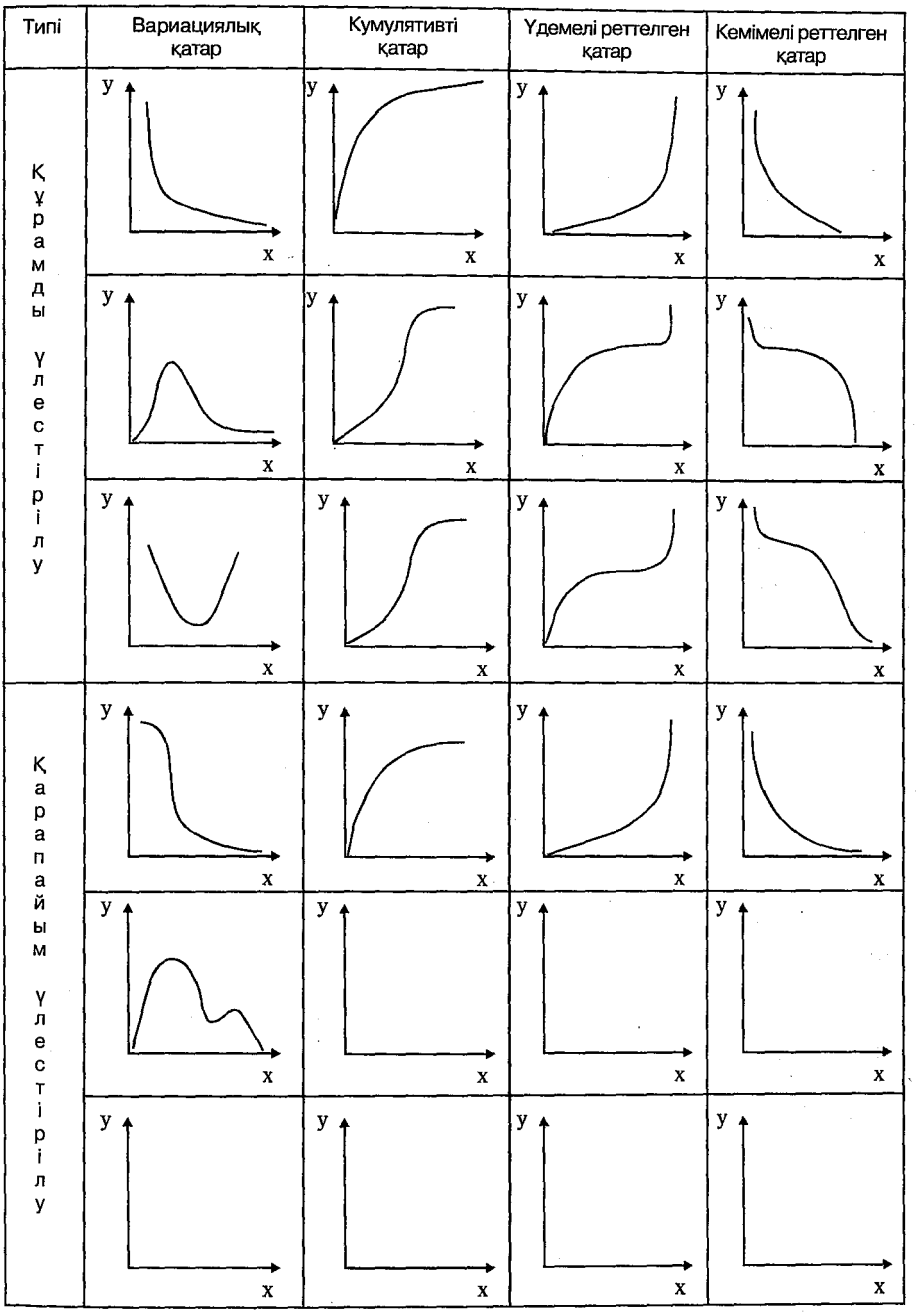
5.3.4. Ранг және жиілік ұғымдары

Ранг пен **жиілік** терминіне анықтама беру үшін белгілі көлемді мәтін бойынша құрастырылған жиілік сөздікті мысал ретінде келтіруге болады. Егер мәтінде кездесетін сөздер (сөзтұлғалар) жиіліктерінің кему тәртібімен реттеліп, ал жиіліктері бірдей сөздер әліпби тәртібімен орналасып, бірінші орналасқан ең жиі кездескен сөзден ең соңғы сирек кездескен сөзге дейін рет саны қойылып (нөмірленіп) шықса, ондай сөздікті жиілік сөздік деп атайтынымыз белгілі. Міне, осындай қатарда екі айнымалы шама бар. Оның біреуі – сөздердің ретін белгілейтін сан, ал екіншісі, оған сәйкес келіп отыратын сөздің жиілік мөлшерін білдіретін сандық шама. Статистика саласында рет санын **"ранг"** деп, ал сөздің мәтінде қайталанып келу санын – **жиілік** деп атайды. Жиілік сөздік мысалында, **"ранг"** өскен сайын, **"жиілік"** кемиді (не тұрақты болады). Яғни екеуінде де екі шама үнемі өзгеріске ұшырап отырады, оларды кездейсоқ шама деуге болмайды.

1-кесте. Кездейсоқ шама мен ықтималдықтың арсындағы функционалдық байланыстың берілу тәсілдері

Теоретика-ықтималдық термин	Сипаттау	Математика-статистикалық термин
Үлестірілулер функциясы $F(x)$	X – кездейсоқ шаманың *x-тен* кіші не тең болуының ықтималдығы – $F(x)$: $F(x)=Prob[X \leq x]=\alpha$; $F(x) = \int_{-\infty}^{x} f(u)\,du$	Кумулятивті қатар
$f(x)$ ықтималдықтың тығыздығы	$f(x)$ функциядан алынған x_L мен x_u аралығындағы анықталған интеграл X кездейсоқ шаманың осы аралықтағы ықтималдығына тең: $\int_{x_L}^{x_u} f(x) = Prob[x_L \leq X \leq x_u]$, $f(x)=\frac{d[F(x)]}{dx}$.	Вариациялық қатар
Үлестірілудің кері функциясы (квантилдердің функциясы) $G(x)$	$G(x)$ –кездейсоқ шаманың қабылдайтын мәні α -дан үлкен мәнге ие болмаудың ықтималдығы $G(x)$-ке тең болатын сан. Яғни: $X=G(\alpha)=G[F(x)]$, $Prob[X< G(x)]=\alpha$	Үдемелі реттелген қатар

Әрбір теориялық үлестірілу, осы тәріздес, кездейсоқ шамалардың мүмкін болатын мәні мен оған сәйкес келетін ықтималдықтың арасындағы тәуелділікті анықтайтын функция түрінде көрініс табады. Мұндай тәуелділікті функция түрінде жазудың бірнеше тәсілдері бар. Солардың ең маңыздылары *1-кестеде* беріледі, ал *2-кестеде* үлестірілудің екі түрі көрсетілді: Ципфа–Парето және көрсеткіштік үлестірілулер. Екі кестеде де Н.Хастинг және Дж. Пикоктың (1980) жүйелеу түрінде көрініс тапты.



1-сурет

1-кестедегі үлестірілулер әртүрлі тәсілмен өзара бірмәнді қарсы мағынада бейнеленген: интегралды тұрпат – жіктемелі (дифференциалды) тұрпат, тура функция – кері функция, оқиғаның ықтималдығы – қарама-қарсы оқиғаның ықтималдығы.

Лингвостатистикалық зерттеулерде, әдетте, бір үлестірілулердің әртүрлі тұрпаттағы берілістерінің арасындағы байланыс дұрыс қабылдана бермейді. Мысалы, рангілер үлестірілулердің екі түрінің (үдемелі және кемімелі) немесе функциялық үлестірілу мен рангтық үлестірілу араларындағы байланыс ескеріле бермейді.

2-кестеде спектрлік пен кемімелі рангтік үлестірілулер араларында ешбір айырмашылық жоқ екендігі көрсетіледі.

2-кесте. Көрсеткіштік пен Ципфа – Парето үлестірілулердің функционалды берілу пішіні

Үлестірілулердің берілу пішіні	Ципфа – Парето үлестірілуі	Көрсеткіштік үлестірілу
Кумулятивті қатар	$K\text{-}K\,x^{-c}$	$K\text{-}K\,e^{-\lambda x}$
Вариациялық қатар	$Kcx^{-(c+1)}$	$K\lambda\, e^{-\lambda x}$
Үдемелі реттелген қатар	$(\frac{K}{K-r})^{\frac{1}{c}}$	$\frac{1}{\lambda} ln(\frac{K}{K-r})$
Декумулятивті қатар	$K\,x^{-c}$	$K\lambda\, e^{-\lambda x}$
Кемімелі реттелген қатар	$(\frac{K}{r})^{\frac{1}{c}}$	$\frac{1}{\lambda} ln\frac{K}{r}$
K – жиынның көлемі, x – құбылмалы белгінің шамасы, r - ранг		

Ал үлестірілулердің негізгі геометриялық типтерінің қарапайым түрі мен құрама түрлері *1-суретте* көрініс тапты.

5.3.5 Әртүрлілік және оны шектеу

Мәтінді туындату екі түрлі үрдіс бойынша іске асады:

а) әртүрлі бірліктер жиынын кеңейтуге немесе әртүрлілікті көбейтуге талпыну жолы;

ә) кейбір бірліктердің жиілігін өсіруге немесе әртүрлілікті шектеуге талпыну жолы.

Мұндай бір-біріне қарсы тұру үрдісі тілдік бірліктердің екі типті сандық вариативтілігіне (**белгі-типтердің** номенклатуралық құбылуы) сәйкес келеді: құрамның вариативтілігі және рекуренттік вариативтілігі (әрбір **белгі-типінің** және барлық белгі типтерін қоса есептегендегі **белгі-қолданыс** санының құбылуы). **Белгі-типтердің** жинақталған сандарымен **белгі-қолданудың** жинақталған сандарының аралығын байланыста ұстайтын үлестірілулерді **белгі-тип– белгі-қолданыс үлестірілуі** деп атайды. Статистикалық лексикографияда мұндай үлестірілу түрі лексикалық әртүрлілікті және мәтіндердің лексикалық шоғырлану сипатын (мәтінді қамтуды) анықтауда пайдаланылады.

Белгі-тип (знакотип) үлестірілуін құруда мәтін бірліктерінің рекуренттік вариативтілік ескерілмесе, ***белгі-қолдану*** үлестірілулерін құруда вариативтіліктің екі түрі де ескерілді. Кейде мұндай сәйкес үлестірілулерді – **сөздіктегі үлестірілулер** және **мәтіндегі үлестірілулер** деп те атайды.

5.3.6 Құрылым және әрекеттік сипат

Танымдық нысанды немесе таным қызметін бір нәрсенің бағытталған екі жақтық қарама-қарсылықтар тұтастығы деуге болады. Оған жататындар: әрекеттік сипат және оның материалдық негізі – құрылымы.

Осыған байланысты статистикалық қатарларды құруда түрленуші белгілердің екі категориясы пайдаланылуы мүмкін:

1) жиынтық бірліктердің ішкі құрылысын қамтып көрсететін белгілер;

2) жиынтықты құрастырушы бірліктердің мәртебесін (функционалды салмағын) білдіретін белгілер.

Аталған белгілердің категорияларына қатысты статистикалық қатарлардың екі түрлі типін атауға болады: **статустық** және **құрылымдық қатарлар**. Мәтін ішіндегі сөзтұлғалар, мысалы, олардың әріп санымен, фонемалар немесе морфемалар санымен, яғни сөзтұлғалардың ішкі құрылымын қамтып көрсететін белгілермен өлшенеді. Бұл жағдайда сөзтұлғалар "ұйымы" құрылымдық үлестірілуді құрайды. Бірақ сол сөзтұлғалар жиынтығы басқаша түрлі өзгеруі де мүмкін. Мысалы, олардың мәтін ішінде кездесу жиілігі бойынша, яғни бұл жағдайда – функционалдық белгілері негізінде реттеледі және тек статустық үлестірілу жайлы ғана сөз болуы мүмкін. Үлестірілулердің екі типінің арасында айқын жатқан шекара жоқ: нақты нысанның функциясы оның құрамымен және, керісінше, нысанның ішкі құрылымы оның функциясымен анықталады. **Статустық** пен **құрылымдық үлестірілулер** арасындағы байланыс жиынтықтардың кез келген табиғатына тән – органикалық емес табиғатынан бастап, таңбалы-ақпараттыққа дейін. Мысалы, әр ортадағы нақты химиялық элементтің салмақтық құрамы оның атомдық салмағына байланысты, ал мәтін ішіндегі сөздің жиілігі – мәтін көлеміне, сөздердің көпмағыналығына, ұғымдар көлеміне және т.б. байланысты болып келеді.

5.3.7 Элемент және олардың жиынтығы

Жиынтық нысандарды статистикалық сипаттау олардың әрбіреуін жеке сипаттау мен жеке нысанға даралауға қажет етпейтін жалпы қасиеттерді жиынтықты сипаттаулар арасында аралық орын алады. Статистикалық әдістерге қызығушылық жеке немесе жиынтық нысандардың қасиеттерін зерттеуге ауысып отыруы мүмкін. Бірінші жағдайда, ғалым-тілшіні қызықтыратын жайт – бір нысанның әртүрлі жағдайларға қатысты әрекеттері (бірнысандық зерттеу). Екіншіде – өзара жақын жатқан әр түрлі нысандардың бір ғана немесе біртекті жағдайлар топтамасы ішіндегі әрекеттік мүмкіндігі (көпнысандық зерттеу). Зерттеу мақсатының айырым табуы екі қарама-қарсы үлестірілулер құруға әкеп соғады. Бірінші типті үлестірілулерді **бірнысандық**, ал үлестірілулердің екінші типін **көпнысандық** деп атауға болады. Бұл үлестірілулердің бір-біріне принциптік ұқсамау жағдаятын ең алғаш байқаған ғылыми статистиканың негізін қалаушы ірі ғалым А.Кетле деп саналады. Оның байқауынша, бір ғана нысанның (мысалы, тек бір ғана үйдің) бір белгісін (биіктігін) бірнеше рет өлшеу нәтижелері мен бір типті бірнеше нысандардың бір ғана белгісін (мысалы, бір көше бойындағы бірнеше үйлердің биіктігі) бірнеше рет өлшеу нәтижелерінің орта шамаларының арасында қатты айырмашылық болады.

Егер тәжірибе нәтижелері екіөлшемді кестелер арқылы берілетін болса, **бірнысанды** және **көпнысанды** үлестірілулер арасындағы айырмашылықтар айрықша түрде байқалады. Оған мысал ретінде ғалым Андрей Белыйдың Гоголь шығармаларындағы түр-түс спектрын қолдану мүмкіндігін анықтауға арналған зерттеу нәтижелерін атап өтуге болады (*3-кесте*).

3-кесте. Гоголь шығармаларындағы түр-түс спектры (А.Белыйдың деректері)

Түр-түс аттары	Шығармалар тобы				Статистикалық орта, %
	I	II	III	IV	
Красное	26,6	12,5	10,3	6,4	17,4
Белое	9,5	9,0	22,0	17,0	14,0
Черное	11,0	14,1	11,8	4,8	12,0
Зеленое	8,6	7,7	9,6	21,6	9,4
Золотое	11,6	8,9	2,8	12,8	9,2
Синее	10,7	6,1	4,9	6,4	8,7
Желтое	3,5	8,5	10,3	12,8	7,0
Серое	2,6	8,9	10,5	6,4	5,8
Голубое	4,4	5,7	7,0	1,6	4,8
Серебренное	7,1	3,2	2,8	4,8	4,8
Коричневое .	0,9	6,5	8,4	1,6	4,0
Розовое	3,8	0,8	2,1	3,2	2,3
Оранжевое ...	0,3	1,6	2,8	0,0	1,2
Лиловое	1,7	1,2	0,0	1,6	0,9

Белгілеулер: I – "Вечера на хуторе близ Диканьки", "Вий" и "Тарас Бульба"; II – повести и комедии; III – первый том "Мертвых душ"; IV – второй том "Мертвых душ".

3-кестенің көлденең жолында берілген зерттеу нәтижелері **бірнысанды** үлестірулер, ал кесте бағаналары бойында орналасқандарды **көпнысанды үлестірулер** деп қарастыруға болады. Дәлірек айтсақ, мұндағы **бірнысанды үлестірілу** деп отырғанымыз – әр топ шығармаларындағы нақты түр-түстің вариациясы немесе құбылуы. Ал **көпнысанды** үлестірілулер дегеніміз – әр топ шығармаларындағы түр-түс гаммасының үйлесімділігі. Осындай кеңістіктік бағытта болу терминология саласында өз орнын дұрыс тапты деуге болады. Әсіресе мұның әсері статистикалық лексикографияда айқын білінеді. Бұл жаңа салада **бірнысанды** үлестірілулерді **"горизонтальды"**, ал **көпнысанды – "вертикальды" үлестірілулер** деп те атап жүр.

Нақты "нысанның" үлестірілуі әртүрлі ортақтық дәрежеде алынуы мүмкін. Мысалы, ол ең қарапайым түрде жиілік сөздік бірліктері немесе барынша жалпылық сипатта болатын – Андрей Белыйдың құрастырған нысандары. Ал бұл, статистикалық тәжірибе нәтижелерінің ортақтық сипаттағы нысандар құру негізінде жатқан статистикалық тұжырмнан тыс тұратын сезгіштік, тапқырлық және тіпті зерттеушінің батылдық қасиеттерінің күшті әсері деуге болады.

5.3.8 Виртуалдылық және актуалдылық

Лингвостатистикада танымдық қызметтің деңгейіне қарай, екі типті жиынтық қарастырылады. Олар – тікелей байқауға қатысты және объектілеу өрісіне қатысты деңгейлер. Бірінші жағдайда мәтіндегі лингвистикалық бірліктердің вариативтілігі зерттеледі де, екіншіде – тілші қызметінің реттелген, жүйеленген және қорытындыланған нәтижелері зерттеледі. Оған жататындар тілші қызметінің терминологиялық және түсіндірме сөздіктегі семантикалық желілеріндегі, ақпараттық-іздеу тезаурустарындағы және т.б. нәтижелер. Объектіленген бірліктер материалдарында (виртуалдылар) құрылған үлестірілулер белгіленген ортадағы

(өрістегі) үлестірілулер деп, ал мәтіндер материалындағы (актуалды) бірліктер негізінде құрылған үлестірілулерді функционалдау аясындағы (өрісіндегі) үлестірілулер деп аталады.

5.3.9 Біркелкілік және біркелкі еместік

Математикалық-статистиканың классикалық көзқарасы бойынша, статистикалық тұрғыдағы жиынтықты анықтайтын негізгі шарт – оның бірліктерінің нақты белгісіне қатысты **біркелкі болып келуі**. Бұл жердегі **біркелкілік** ұғымы тек сапалық сипатқа ғана қатысты емес, сонымен бірге сандық сипатты да қамтиды. Егер жиынтықтың біркелкілігі сандық белгі бойынша анықталса, онда ол **сандық біркекілік** деп, ал сапалық белгі бойынша – **сапалық біркелкілік** деп аталады.

Біркелкіліктің шарты ретінде статистика саласында сандық белгілер мағынасының бір-біріне жақын тұру сипаты алынады және ол жақындық критерийі деп аталады. Мұндай жағдайда сандық белгі өзі дара тұрмай, құбылыстың бірсапалық шартымен бірге тұрып қабылданады. Яғни сандық біркелкілік белгілі мағынаға ие болуы үшін ол сан мен сапалық тұтастықта болуы қажет. Мұны біркелкіліктің өлшемі деуге де болады. Бұл шарттан ауытқу деген сөз – жинақталған сандық және сапалық өзгерістер негізінде бір сападан екінші сапаға ауысуға мәжбүр болу.

Осы күнге дейін жиынның сандық дәрежеде біркелкі болуының негізгі критерийі нормальдылық критерийі болып келді. Бұл критерийді өткен ғасырдың 20-жылдарында белгілі статистик Б.С.Ястремский ұсынған болатын. Ол былай деп жазған: "Варианты признака однородной совокупности обладают общим уровнем отклонения, от которого оказываются случайными: ни один из множества факторов не выделяются своим влиянем на общий результат. Именно такой характер влечет за собой в ряде случаев образование нормального распределения аналогично закону распределения прогрешности.

Варианты признака разнародной совокупности лишены общего уровня и вследствие этого, не укладываются в нормальное распределение. Наличие асимметричного распределения сигнализирует о необходимости тщательных поисков возможных источников разнородности" [Ястремский, 1960, 92-б.].

Б.С.Ястремскийдің қолданған әдісін жетілдіре түскен В.И.Сиськов (өзінше бөлек) сандық жағынан біркелкі болып келетін топтарды құрудың алгоритмін ойлап тапты. Біркелкілік критерийі ретінде ол вариация коэффициентін ұсынады: егер топтама ішінде вариация коэффициентінің мәні 0,33 мәнінен кіші болса, ондай топ біркелкі деп саналады [Сисков, 1975].

Көптеген жағдайларда біркелкі жиынды қалыптастыру тәжірибе жүргізу арқылы іске асады. Яғни жиынды сапалық талдау арқылы **біркелкі емес жиын сапалы біркелкі** топтарға бөліктенеді. Осындай талдаудың негізінде бірліктердің сапалық жағынан алғандағы **біркелкі** болып келетін бірнеше үлестірілулер құрылады. Бірақ көптеген жағдайда, лингвистикалық нысандардың сапалық ерекшелігінің негізінде нақты нысанның берілген "ішкі жиындарға" қатыстылығын тікелей танып-білу мүмкін болмай шығады. Бұл жиынды біркелкі емес деп тұжырымдаудың дәлелсіз, тек қана жорамал екендігінің айғағы. Яғни біркелкі еместікті жою априорлы деген сөз. Осы айтылғанның негізінде үлестірілулерді **априорлық жойылатын** және **априорлық жойылмайтын біркелкі еместік**, деп екіге ажыратуға мүмкіндік туады. Лингвистикалық бірліктердің ішкі жиындар арасында да анық түрдегі шекараның

болмауынан, **априорлық жойылмайтын біркелкілік** пайда болады. Ондай шекараларды айқындау тек қана статистикалық топтау теориясындағы жете зерттеулер негізінде ғана шартты түрдегі формалды әрекеттер арқылы іске асады.

Кейде **біркелкілік** пен **біркелкі еместік** сипат, топтарды құрайтын біркелкіліктің ұқсастығы арқылы түсіндірілмей, олар әртүрлі топтар үлестірілулерінің жақындығы тұрғысынан дәлелденеді. Бұндай жағдайда үйлесімдік критерийі негізінде қалыпты сызбалар арқылы екі таңдама бөліктердің (топтардың) тек бір ғана бас жиынға қатыстылығы тексеріледі.

5.3.10 Орнықтылық және орнықсыздық

Статистикалық шамалардың орнықтылығы жайлы мәселе, статистердің көп жылдар бойы бірізді шешімі табылмай келе жатқан мәселесі болып саналады. Бүгінгі күннің өзінде кейбір ғалымдар бұл мәселеге байланысты өз ойларын былай білдіріп жүр: "Вопрос о причинах устойчивости частот можно считать только поставленным. Его решение еще не предвидится, хотя некоторые усилия в этом направлении предпринимаются" [Сачков, 1971, с.123].

Бұл проблеманың шешімінің оңай табылмауының негізгі себебі – ықтималдықтар теориясының айналысатын сызбаларына тән, абсолюттік орнықтылықтың ақиқат шындық құбылыстары ішінде кездеспеуі. Статистикада, оның ішінде лингвистикалық статистикада да статистикалық қатарлардың, кеңістік пен уақытқа қарағанда аз өзгеріске ұшырайтын, тек салыстырмалылық орнықтылық жайлы сөз ету ғана дағдыға айналған. Бұл, эмпирикалық ақиқаттықты ықтималдықтар терминімен түсінік беруде белгілі сақтылықты қажет етеді. Себебі ықтималдықтар теориясының қолдану аясында кездесетін құбылыстар жиілігінің орнықтылық сипаты дәлелдеуді қажет етпейтін аксиома ретінде түсіндіріледі. Сонда да, бұл аксиома, тәжірибеші-статистиктерді ақиқаттың жаппай процестерінің заңдылықтарын ықтималдықтар теориясы негізінде танып-білуге бағытталған іс-әрекетіне кедергі бола алмайды. Әрине, бұл әрекеттер де негізсіз емес.

Бұл жерде жоғарғы қазының (арбитр) рөлін атқаратын, теорияны тәжірибемен "өлшейтін", әрі мән-мағынасы математикамен шектелмейтін – үлкен сандар заңдылығы орын алады. Аталған заңдылық – сырттай статистикалық сандардың орнықтылығын бейнелейтін ақиқат шындықтың заңы. Осыған қатысты А.А.Чупров былай деп жазған: "Устойчивость статистических чисел, их свойство начинает колебаться из года к году, лишь в известных, ограниченных пределах представляет собой эмпирически установленный факт, который сам по себе, независимо от тех или иных теоретических истолкований, имеет громадную научную и жизненную важность. Это один из коренных, хоть и мало заметных, устоев современной культуры" [Чупров,1910].

Кез келген жете дайындықпен ұйымдастырылған эксперимент теоретика-ықтималдық сызбасыз-ақ, өзінше бір бөлек, практикалық құндылыққа ие болуы мүмкін. Бірақ тағы да А.А.Чупровтың сөзімен айтсақ, "лишь на почве теории вероятностей удается осмыслить наблюдаемые в статистических числах колебания, лишь теория вероятностей открывает возможность прагматического их истолкования в каждом отдельном случае" [Чупров, 1910].

Аталған факторлар эмпирикалық үлестірілулердің орнықтылық критерийін табуға негіз болады.

Бұл бағыттағы алғашқы қадам жасаған неміс статисті В.Лексис болатын. Ол статистикалық қатарлардың орындылығын өзгерту үшін **дивергенция коэффициенті** (**Q²**) деп аталатын ерекше көрсеткішті ұсынды [Lexis, 1903]. Бұл көрсеткіштің мағынасы мынада. Эмпирикалық дисперция теориялық шамамен салыстырылады. Ал теориялық шаманы табу үшін барлық статистикалық қатардағы қатынастық жиілікті шартты түрде ықтималдыққа тең деп қабылдайды:

$$Q^2 = \frac{2\sum(p_i - p)^2}{m-1} : \frac{2p(1-p)}{n}$$

Мұндағы m – әрбір "n" байқаулар ішіндегі сериялар саны, P_i – сериялардағы қатынастық жиілік, P – ықтималдыққа тең жалпы қатынастық жиілік (общая частнось).

Лексис бойынша, егер Q^2 шамасының *"1"* санынан барынша аз мөлшерде ауытқыса, онда үлестірілу – нормальды деп, ал егер Q^2 шамасы *"1"* санынан кіші болса – қатардың орнықтылығы нормальдықтан жоғары және ол шама *"1"* санынан көп болса, қатардың орнықтылығы нормальды жағдайдан төмен деп саналады.

Сонымен Q^2 шамасына сай, статистикалық қатарлардың орнықтылығы жайлы нормальды сипатта болуы жайлы немесе орнықтылықтың нормальдылықтан төмен және жоғары болуы туралы сөз етуге болады.

Лексистің зерттеу нысаны ретінде қарастырған аса "ұзын" сериялы статистикалық қатарлардың ешбіреуінде де орнықтылықтың нормальдылықтан жоғары жағдайы орын алмаған. Яғни Гаусс заңдылығының қателеріне тән, орнықтылық дәрежесі Лексис бойынша ең үлкен (максимум) мәніне тең. Басым көпшілік жағдайда эмпирикалық қатарлардың орнықтылығы нормадан бес, он тіпті жиырма реттен төмен болып шыққан.

Лексистің және оның көптеген шәкірттерінің эксперименттік нәтижелерін ғалым А.А.Чупров қайта қарастырып, ақиқат өмірдің құбылыстарында статистикалық қатарлардың орнықтылықтарының нормальдыдан жоғары болу сипаты айтарлықтай жиі кездесетіндігін дәлелдеді. Ғалым, аталған фактіні тілдік материалдар негізінде, дәлірек айтсақ, Гетенің шығармалары мәтінінен алынған бірнеше таңдама сериялардағы әріптердің үлестірілулерінің негізінде көрсетіп дәлелдей алды.

Қазіргі кезде тұрақтылық көрсеткіші ретінде әртүрлі критерийлер қолданылып жүр. Солардың ішінде тілдік материалдарды зерттеуде жиі қолданыс тауып жүрген X^2 (хи-квадрат) критерийін атауға болады. X^2 критерийін қолдануға байланысты Б.Н.Головин былай дейді: "...Полезно было бы различать три степени статистичекой устойчивости явлений языка и речи: первая (оптимальная устойчивость) – показатель "хи-квадрат" не выходит за пределы границы существенности, вторая (нарушаемая устойчивость) – величина "хи-квадрат" превосходит границы существенности более чем в два раза" [Головин, 1971, с. 90-91].

Дәстүрлі вариация коэффициенті мен оның жетілген түрлері, мысалы, Жийянның тұрақтылық коэффициенті сирек қолданыс тауып жүргенін айта кетуге болады. Бұларға қарағанда, жиілік өлшемі мен олардың көп таралуына негізделген көрсеткіштер тіпті де сирек қолданылуда [Андреев, 1965; Кобрин, 1976; Марусенко, 1980].

5.3.11 Сиректік және жиілік

Классикалық статистикада жиі кездесетін элементтермен әрекет ету **нормальды заңдылық** көмегімен және оның негізін құраушы **шекті теоремалары** арқылы іске асып жүр. Ал сирек кездесетін оқиғалар **Пуассон заңдылығы** немесе **сирек оқиғалар заңдылығы** деп аталатын заңдылықтар арқылы зерттеледі. Бұлайша әрекет ету, ақиқат процестерді модельдеу кезінде де және үлестірілудің белгісіз өлшемдерін бағалауда да бірдей дәрежеде қолданылып жүр.

Тіл білімінде тілдік элементтердің сирек не жиі үлестірілулері арасындағы айырмашылықты былайша ұғынуға болады: мәтін ішіндегі жиі кездесетін семантикасы жойылған, доминанттық емес элементтер – нормаль заңдылығына бағынады, ал семантикалық деректі, доминанттық сирек кездесетін тілдік бірліктер – Пуассон заңына бағынады. Көркем шығармаларды сюжет пен композиция тұрғысынан алып зерттеуде семантикалық доминанттықтар ретінде сөздер мен сөз тіркестер алынып жүр. Мұндай фактіні проф. Р.Г.Пиотровский былай түсіндіреді: "... Порождение текста определяется, с одной стороны, системой языка и его нормой, а с другой – независимой от языка ситуацией, которую призван описывать этот текст. Появление семантически нагруженных ключевых элементов текста диктуется ситуацией, употребление же семантически "стертых" единиц, так называемых единиц заполнения, служащих средством организации текста и связи ключевых единиц, подчиняется требованиям системы и нормы языка" [Пиотровский, Бектаев, Пиотровская, 1977, с. 356-357].

Мәтінде қолданыс табатын жиі және сирек элементтер, олардың толық жүйедегі рөліне қарай да және артықшылығы жағынан да зерттеушілердің ынтасын аударуда [Арапов, Шрейдер, 1970]. Аталған жайт стилеметрия мәселесінде де аса маңызды деп саналады [Орлов, 1970; Борода, 1977].

5.3.12 Симметрия және асимметрия

Лингвистикада симметриялық үлестірілулердің сирек кездесетіні белгілі. Көпшілік жағдайда, зерттеушілер бөсең және өте-мөте асимметриялық сипаттағы үлестірілулермен әрекет етеді (*1-суретті* қара). Асимметрия дәрежесінің әртүрлі болуы мынадай жағдайларға қатысты болады:

1) бір жиынға, сапалық жағынан алғанда, біркелкі емес элементтердің немесе олардың топтарының біріктірілуі;

2) жиын бірліктері арасындағы өзара байланыстың болуы;

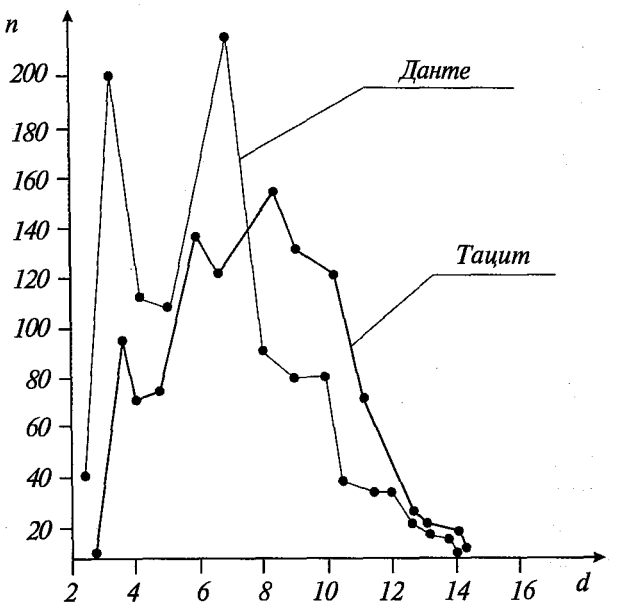
3) құбылмалы белгінің бастапқы немесе шеткі мәндерінің қайсыбір санмен шектеулі болуы;

4) лингвистикалық бірліктің нақты тілдік деңгейдің немесе лексика-грамматикалық кластың құрамында болуы және т.б.

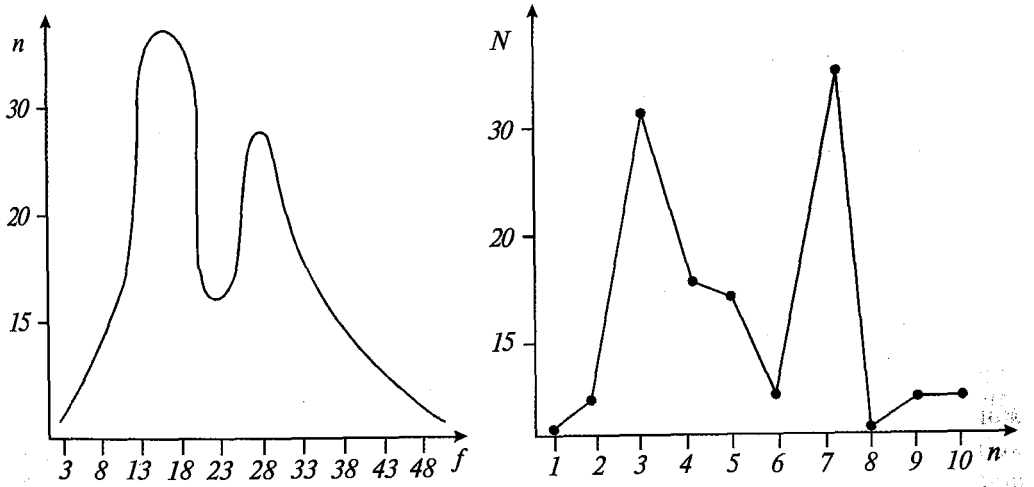
5.3.13 Біртөбелік және көптөбелік

Лингва-статистикалық зерттеулерде біртөбелік (бірмодальды) вариациялық қатарлар мен көптөбелік (полимодальды) қатарлар кездеседі. Вариациялық қатардың көптөбелік болуы қатарды құрайтын материалдың әрдайым біркелкі болмауына

нұсқайды. Дей тұра, кейбір құбылыстардың үлестірілулеріне олардың құрылымдық ерекшелігіне қарай полимодальдық сипат та тән. Ең алғаш, осындай тілдік құбылыстарға көңіл аударған белгілі орыс статистигі П.А.Вихляев болатын. Ғалым Тациттың, Цезарьдың, Дантенің және Леопардидың шығармаларындағы сөзқолданыстарды тәжірибе ретінде қарастырады. Егер олардың ұзындық өлшемін құбылмалы белгі ретінде алып, үлестірілулерін құрастыратын болсақ, олар әрдайым екітөбелі болады екен.



2-сурет. Үлестірілулердің негізгі геометриялық типтерінің қарапайым және құрама түрлері



3-сурет. Екітөбелік қатарлардың геометриялық пішіні

П.А.Вихляев оның себебін былай түсіндіреді: "Причина этого явления кроется в самом свойстве языка, который слагается из коротких частей – предлогов и союзов и более длинных – прилагательных и существительных" [Вихляев, 1928].

Үлестірілулердің көптөбелігін көбінде "трансгрессия" құбылысымен түсіндіреді. Оның мәні мынада: екі жиынның бір-біріне "құйылуында" біреуінің белгі мәнінің үлестіру максимумы екіншісінің минимум шамасымен беттеседі. Егер мұндай вариациялық қатарлардың арифметикалық орталарының айырмашылығы статистикалық тұрғыда ақиқаттыққа жатса, онда бұларды трансгрессияланатын қатар деп атайды. Екітөбелік қатарлардың геометриялық пішіні *2-ші* және *3-суреттерде* көрсетілді.

Бұл сызбалар, жоғарыда аты аталған ғалым П.А.Вихляевтың сөзқолданыстардың "ұзындығы" бойынша түзілген үлестірілулерінің деректері бойынша сызылған. Мұндағы d – сөзқолданыстың ұзындық өлшемі, n – оның жиілігі.

Лингвистикалық қатарлардың полимодальдығы кейде жасырын сипатта да болады. Мұндай сипат әдетте үлкен қарқынды құбылмалы вариациялық қатарлардың үлестірілулерінде пайда болып, қатты асимметриялық әсер қалдырады. Кейбір үлестірілулерге тән мұндай сипат таңдама бөлік көлемінің шағын болуынан деп те түсіндіріледі.

5.3.14 Статикалық және динамикалық қатарлар

Құбылыстардың уақытқа қатысты өзгеріске ұшырауын зерттеу статистика пәнінің ең маңызды мәселелерінің бірі деп саналады. Аталған мәселе динамиканың қатарларын (динамикалық немесе уақытқа қатысты қатарлар) құру мен талдау арқылы шешіледі. Динамикалық қатар дегеніміз сандық қатар, ал оған жататындар – уақыт тізбегінің немесе кезендік уақыттардың статистикалық көрсеткіштерінің нақты мәні. Мұндай көрсеткіштер мәні сандық қатардың деңгейі деп те аталады. Егер статистикалық қатарларда аргумент рөлін жиын бірліктерінің ішкі не сыртқы құрылымын сипаттайтын белгілер атқарса, динамикалық қатарларда ондай аргумент ретінде уақыт категориясының әртүрлі түсінігіндегі уақыттық сипаттамалар атқарады.

"Уақыт" категориясына қандай мән-мағына берілуіне қарай, филологиялық зерттеулерде қолданыс табатын динамикалық қатарлардың бірнеше түрлерін ажыратуға болады. Олар статика және динамика, синхрония және диахрония деп аталатын екі жазықтықта айырым табады.

Айырым белгі, яғни динамиканың статикадан айырым табуы – сөйлеуді (мәтінді) туындатуды динамикалық процесс арқылы түсінік беруге талпыну. Динамика "тірі" туындайтын процесс ретінде түсінік алуы мүмкін немесе қайсыбір ойды тіркеу мен талдау мақсатындағы тұрақты процесс деп те қарастырылады. Соңғы жағдайда зерттеушінің ынтасы, негізінен, мәтіннің сызықтық және уақыттық құбылуына бөлінеді. Бұл жерде мәтін бір-біріне тізбектеліп, тіркесіп жатқан кесінділер тәріздес түсінік алады. Ал әрбір кесіндіде кездейсоқ айнымалы шаманың нақты мағынасы тіркеледі деп ұғыну қажет.

Мәтін құрылымы – сюжеттік, композициялық, тақырыптық, "кейіпкерлік" және т.б. бөліктердің тізбегі ретінде түсінік алуы мүмкін. Яғни бұл жағдайда, көркемдіктің уақыттық сипаты орын алады деуге болады. Мәтіннің қай түсінігін арқау етсек те (тіл білімі немесе әдебиеттану ғылымы), біз әрдайым ішкімәтіндік уақытқа сүйенеміз. Мұндай жағдайды негіз етіп құрылатын мәтіндерді **сызықтық** және **уақытқа қатысты мәтіндер** деп атайды.

Динамикалық қатарлардың айырым табатын екінші белгісі – синхрония мен диахронияны бір-біріне қарсы қою. Синхрониялық қатарларға қарағанда диахрониялық қатарларда тарихи және эволюциялық факторлар ескеріліп отырады. Стилеметиялық зерттеулерде жеке-стильдік, жанрлық, тектік және басқа да әдеби жүйелердің эволюциясын (дамуын) анықтап-білу қарастырылады. Мұндай жағдайда уақыт ұғымы тарихи, дәуірлік, календарлық және т.б. мағынаға ие болуы мүмкін.

Қорыта айтсақ, үлестірілулердің типтерін ажыратуға негіз болатын белгілер лингвистикалық және статистикалық ақиқаттықтың әртүрлі жақтарына қатысы бар. Олардың бір бөлігі теоретика-ықтималдық түсінікке, келесі бөлігі эмпирикалық деректердің берілу тәсілі мен олардың көрінісіне, ал үшіншілері зерттеу жиынының ерекшелігіне, бірліктеріне және құбылмалы белгісіне негізделеді. Жоғарыда бөлектеліп қарастырылған үлестірілулер түрлері бірдей "жүйелік салмаққа" ие емес, бірақ олардың әрбіреуінің зерттеу мақсатына сай маңыздылығы арта түсуі мүмкін.

Үлестірілулердің лингва-статистикалық теориясы өзінің қалыптасу сатысында ғана тұр деуге болады. Ал тілшілердің күш-жігері, негізінен, статистика ғылымы мен тіл ғылымының ұғымдық және категориялық аппараттарының үйлесімді дамуына бағытталуы қажет болады.

5.4 Қазақ мәтінінің ықтималды-статистикалық моделін құру мәселесі

Қолданбалы лингвистиканың көптеген маңызды мәселелерінің бірі – мәтін ішіндегі тілдік бірліктердің орналасу моделінің ықтималды-статистикалық әмбебап сызбасын құру. Мұндай күрделі істі автоматты түрге келтіру: ақпаратты іздеу, тілдерді тиімді жолмен үйрету, базалық (негізгі) тілді құру, тілдің "нормалық" табиғатын түсіну, тілді стилистика тұрғысынан зерттеуде және т.б. кешенді мәселелерді шешуде аса қажетті деп саналады [Фрумкина, 1971].

Тілді математикалық модельдеу кезінде тәжірибе жүзінде алынған сандық деректерге статистикалық талдау жұмысы жүргізіледі. Мұндай талдау арқылы кейбір тікелей байқауға жатпайтын бірқатар тілдік жайттардың аражігі анықталады. Ал тілдің ықтималдық моделі көмегімен мәтін ішіндегі кейбір бірліктердің сандық қасиеттерін сипаттауға мүмкіндік туады. Егер осы анықталған бірліктер сипаттамасын ақиқат шындықтың көрсеткіші ретінде қабылдасақ, онда модельден нақты нысанға өту іске асады. Ол үшін зерттеуге алынған мәтін бірліктерінің сандық параметрлерін анықтай білу қажет. Бұл параметрлерге "арифметикалық орта" $\bar{X}$мен "стандартты" *(S)* жатқызуға болады. Әрі қарай $\bar{X}$пен S мәндері ықтималдық модельдің осы тәріздес параметрлерімен салыстырылады. Тілдік бірліктердің сипаттамалары зерттеушіні сол тәжірибеге түскен шағын көлемдегі нақты мәтінге қатысты ғана қызықтырмайды, ол бас мәтін жиынтығы сипаттамасы тұрғысынан қарастырғанды қалайды. Бұл жердегі "бас мәтін жиынтығы" (генеральная совокупность текстов) дегеніміз тілдің толық түрі немесе бөлек бір стиль, тіл қабаттары, жазушы тілі және т.б. Бұлайша қарастырудың себебі аталған "бас мәтін жиынтығын" қамтитын толық түрдегі мәтін бойынша статистикалық талдау жүргізу тәжірибе жүзінде мүмкін бола бермейді. Яғни зерттеуге жататын нысан "бас мәтін жиынтығының" таңдалып алынған жеке бөліктері немесе "таңдама бөліктер". Міне, осындай таңдама мәтін бөліктер өлшемдерін бас мәтін жиынына қатысты статистика-математикалық

зерттеу негізінде ықтималды модельден тілдік нысанның өзіне көшу жүзеге асады. Мұндай "көшу" процесі нақты критерийлерге сүйеніп, статистика-лингвистикалық болжамды тексеру негізінде бірліктердің мәтін ішінде кездейсоқ оқиға ретінде "үлестірілу" түрпатын анықтау арқылы іске асады.

Мәселен, мынадай статистика-лингвистикалық болжамды сынау қажет болды делік: мәтін ішіндегі зат есімге қатысты сөздер математикалық статистиканың "нормальды үлестірілу" деп аталатын теориялық заңдылығына бағынады. Мұндай болжамды тексеру үшін зат есім сөздердің эмпирикалық үлестірілу деректерінің теориялық үлестірілу заңдылығы – "нормальды" заңдылығымен сәйкестік дәрежесін тексеру қажет. Ал ондай тексеру тек қана зат есім сөздердің таңдама бөліктер ішінде қайталану жиілігінің орта шамасының бірнеше "сериялары" бойынша қарастырылады.

Кездейсоқтық жағдайдағы лингвистикалық шаманың мәтін бойында таралу (үлестірілу) түрпатының не өлшемдерінің сынауға ұсынылып отырған болжамға сәйкестігі нөлдік (негізгі) болжам *(H_0)* немесе оған қарама-қарсы қойылатын альтернативті болжам *(H_1)* ретінде тұжырымдалуы мүмкін. Негізгі – нөлдік *(H_0)* болжам белгілі критерий арқылы сыналады. Критерийлер *статистикалық* және *реттік* деп екіге ажыратылады. Алғашқысы (статистикалық) кездейсоқ лингвистикалық шаманың үлестірілуіне (қолдану жиілігіне) бағыттала отырып, нөлдік болжамды тексеру үшін қолданылады, ал екіншісі – аталған шамалардың реттік үлестірілуін қарастырады.

Әдетте, лингвистикалық болжамдар кездейсоқ таңдама бөліктер негізінде тексеріледі. Бірақ бұл тілдік бірліктің бас жиындағы үлестірілу көрінісінің дұрыстығының айғағы бола алмайды, себебі, мұндай жағдай "болжамды сынау" процесін жалған жолға бұрып жіберуі де мүмкін. Сондықтан дұрыс шешім орнына, мысалы, дұрыс нөлдік *(H_0)* болжам орнына, бұрыс *(H_1)* болжам (не керісінше) қабылданады. Осындай жайттарға байланысты зерттеушінің қай болжамды *(H_0* не *H_1)* негізгі деп қабылдауының көп мәні бар. Осы айтылғаннан жіберілетін қателердің мәнділік дәрежесі екі түрлі дәрежелік мәнмен ажыратылады. Олар – *1-ші* және *2-ші* дәрежелік мәнді қателер деп саналады. Тіл білімінде қатенің мәнділік дәрежелерін таңдау, көбінде, субъективті ой қорыту негізінде іске асады. Мысалы, түркі, моңғол, тунгус-маньчжур тілдерінің "алтайлық гипотеза" деп аталатын генетикалық туыстығы жайлы болжам белгілі статистикалық критерий арқылы тексеруге болатындай тұжырымдалған [Клоусон, 1969]. Мұндағы нөлдік болжам бойынша барлық алтай тілдері бір ғана генетикалық туыстас жеріне, яғни негіз-тілге қарай бағышталады. Осы тұжырымды теріске шығаратын, өзіне нағыз кәміл сенетін алтайшы да, аса қауіпті *1-ші* дәрежелік мәндегі қате жіберген болар еді. Бұл жерде нөлдікке қарсы қабылданған болжамның өзі дұрыс болмаса да, мәнділік жағынан *2-ші* дәрежелі қате қауіпсіздеу деуге болады. Сондықтан болжамды тексеру критерийін таңдау, сонымен бірге сынау аясын таңдаумен ұштасады деуге болады.

Сөйтіп, қандай да болсын дұрыс шешім қабылдар тұсында *2-ші* дәрежелі қате жіберуден гөрі *1-ші* дәрежелі қате жібермеу жолы аса маңызды деп саналады. Қате дәрежесінің мәнділігін азайту мақсатымен болжамдар өзара тексерілуі қажет.

Тәжірибелік құндылығына қарай, яғни *1-ші* дәрежелі қатеден күтілетін салдарына қарай, зерттеушіні қанағаттандыратын маңыздылық деңгейі таңдалады. Басқаша айтсақ, жіберген қатеден туындайтын салдардың сипатына қарай маңыздылық деңгейін де төмендеткен жөн болады. Әдетте, лингвистикалық зерттеулерде маңыздылық деңгейінің $\alpha=0,05$ немесе $\alpha=0,01$ шамасы да жеткілікті деп саналады. Мәселен, алтай тілдеріне қатысты болжамды статистикалық

тәсілмен тексеру үшін көрсетілген маңыздылық деңгей шамасы қанағаттандырарлық деп саналады.

Лингвистикалық статистика қазір өзінің алғашқы кезеңін, яғни лингвистикалық құбылыстарды сипаттауды аяқтап, келесі күрделі кезеңіне өтуде. Бұл – тілдік заңдылықтарды сандық тұрғыда алдын ала болжай алатын теория құру мәселесі.

Сонымен, лингвистикалық бірліктердің мәтін ішінде үлестірілуінің математикалық тұрпатын анықтау мәселесі тілдік құбылыстың моделін құрудағы ең бір маңызды іске айналып отыр [Фрумкина, 1962].

Аталған мәселе бойынша, қазақ тілі мәтіні ішінде сөз таптарының статистикалық үлестірілуінің кейбір теориялық заңдылықтарға сәйкес келу болжамын сынауға қатысты зерттеулер жүргізілді. Теориялық үлестірілу заңдылықтары ретінде математикалық статистикадағы "нормальды үлестірілу", "Пуассон үлестірілуі" және "Шарлье үлестірілуі" заңдылықтары қарастырылды.

Қазақ тілінің мәтіндік нысаны ретінде М.Әуезовтің "Абай жолы" романынан **10 мың сөзқолданыс**, газет мәтінінен **50 мың сөзқолданыс** таңдама бөліктер тәжірибе ретінде алынды. Ал сөз таптары ішінен ең негізгілері деп: ***зат есім, етістік, сын есім*** және ***үстеу*** сөздердің мәтін ішіндегі статистикалық үлестірілуі сынаққа түсті.

Белгілі бір лингвистикалық бірліктің (мысалы, белгілі бір сөз табының) тілдегі қолданылу сипатын танып-білу кезінде құбылу заңдылығы бұзылмайтын ең кіші мәтін көлемін анықтау зерттеу ісінің маңызды шарттарының бірі.

Осындай мақсатпен жүргізілген тәжірибе реті төмендегідей:

1) мәтін ішіндегі сөздерді сөз табына ажырату және оларды шартты белгілермен сәйкестендіру;

2) **мәтіннің жалпы көлемі** ***25, 50, 100, 200*** және ***500*** сөзқолданыстан тұратын **микробөліктерге**, яғни **сериялық таңдама бөліктерге** бөлу;

3) өзгермелі шамаларға қатысты шартты белгілеуді ұстану: әр стиль бойынша алынған мәтіннің жалпы көлемі – N; микробөлік (серия) көлемі – k; әр сарияға сай микробөлік саны – n.

4) әр жағдайдың жалпылама жазылу пішінін қабылдау – $B(N/k)=n$ (теңдік белгісі шартты). Мұндағы $N=nk$.

Мысалы, "Абай жолы" романы мәтіні бойынша алынған ***10000*** сөзқолданыстан тұратын мәтін көлемін үшке бөліп: $N_1=2500$, $N_2=5000$, $N_3=10000$ қарастырсақ, әрбір N_i көлемді $k=25, 50, 100, 200, 500$ сөзқолданыстан тұратын **микробөліктерге** ***(n)*** бөлсек, әр бөлікке сай, олардың саны да өзгеріп отырады. Сол сияқты, газет мәтіні бойынша тәжірибеге алынған ***50000*** **сөзқолданыстан** тұратын мәтін көлемі де осылайша серия бөліктеріне бөлініп зерттелді.

5) әрбір вариантқа қатысты микробөліктер ішіндегі қажетті сөз табының жиілігі саналып, бірдей жиілікпен кездесетін сөз табының серия саны анықталды. Осы деректердің негізінде жиілік пен микробөлік санынан тұратын дискретті вариациялық үлестірілімдік қатар түзілді.

Жүргізілген тәжірибе бойынша, көркем әдебиет пен газет мәтіндеріндегі негізгі сөз таптарының мәтін бойында таралуының **вариациялық қатарларын** (статистикалық үлестірілуін) кейбір теориялық заңдылықтармен үйлесімділік деңгейі зерттелді. Теориялық заңдылықтар ретінде **Пуассонның, Шарльенің (*A, B* түрлері)** және **нормальдық, логарифмды-нормальдық үлестірілу** заңдылықтары алынды.

Аталған теориялық үлестірілімдік заңдылықтардың **Пуссон** және **Шарлье В** түрі дискретті вариациялық қатар құру арқылы, ал **нормальды, логарифмды-нормальды** және **Шарлье А** түрі үздіксіз вариациялық қатар құру негізінде олардың теориялық жиіліктері мен статистикалық параметрлері анықталды.

Сонымен, құрылған дискретті және үздіксіз вариациялық қатарлар мәтінге қатысты сөз таптары жиіліктерінің эмпирикалық үлестірілімдік заңдылығын анықтайды. Арнайы компьютерлік бағдарламалар негізінде теориялық жиілік пен статистикалық параметрлер есептеліп шығарылуы қажет.

Енді, осындай зерттеуді, яғни эмпирикалық үлестірілудің таңдаудан өткен теориялық заңдылықпен сәйкестігінің мәнді не мәнсіз екендігін анықтау үшін жүргізілетін тәжірибе реті төмендегідей [Джубанов, 1987]:

1. Зерттеуге тиісті мәтінді және тілдік бірлікті таңдау. Мысалы, қазақ көркем әдебиет мәтіндеріндегі **зат есім сөз табына қатысты сөздер.**

2. ***B(N/k)=n*** көрінісіндегі таңдалып алынған параметрлеріне ***(N, n, k)*** сай етіп, зерттелетін мәтінді таңдама бөліктерге бөлу.

3. ***N, n, k*** параметріне және тілдік бірлікке сәйкес жиілік сөздіктер түзу.

4. ***Жиілік сөздіктер негізінде дискретті не үздіксіз вариациялық қатар құру.***

5. Тиісті математикалық өрнекті (формуланы) пайдаланып, жоғарыда сөз болған мына параметрлерді есептеп шығару:

а) математикалық күтілуді ***(a);***

ә) ауытқудың квадраттанған ортасын $(\sigma\ )$;

б) асимметрия ***(A)*** мен эксцесті ***(E)***.

в) тілдік бірліктің теориялық жиілігін $(n_i^T = N \cdot P_i\ )$;

г) әрбір ***i*** тобына қатысты "хи квадрат" $(\chi^2\ )$ мәнін.

6. Жиіліктің ***5-тен*** кіші мәндерін қажеттілікке сай біріктіру және $\sum_{i=1}^{k} \chi_i^2$ қосындысына қажетті өзгертулер енгізу.

7. Еркіндік дәреже санын анықтау. Мысалы, үлестірілудің – **Пуассон заңы** үшін $\lambda\ =1$, **нормаль** мен **логнормаль**, **Шарлье В** заңдары үшін $\lambda\ =2$, **Шарлье** *A* үшін $\lambda\ =4$.

8. Есептеліп шығарылған χ^2 пен еркіндік дәреже саны ***(ν)*** бойынша – $p(\chi^2 \geq \chi_q^2)$ ықтималдық интегралдың кестелік мәнін табу [Романовский, 1947].

Лингвистикалық зерттеулерде сериялар саны мәтіндегі сөзқолданыстар санынан басқа кітаптың жол саны не бет санымен де өлшенуі мүмкін. Сол сияқты өлшемдерді "қалыпқа түсіру" (нормалау) кезінде мәтін бірліктерінің қайсысы және қандай көлемде алынатыны алдын ала нақтылануы қажет. Әрбір нормалаудың сипатына сай және өзіне ғана тән теориялық үлестірілу заңдылығы сәйкес келуі мүмкін.

Мәселен, нақты бір нормалау сынаққа алынған тілдік бірлік "Пуассон үлестірілу" заңына, ал басқаша түрдегі нормалауда сол бірлік мәтін ішінде "Нормаль үлестірілу" заңына "бағынуы" немесе белгілі бір ішкі бөлік көлемінен бастап тек бір ғана теориялық заңдылыққа сәйкес келуі мүмкін екені анықталды [Бектаев, Зубов, Ковалевич, 1969, Лукьяненков, 1970, Феллер, 1967]. Сондықтан үлестірілу заңдылықтарының өлшемдерін салыстыруда және оған лингвистикалық тұрғыда түсініктеме беруде зерттеу нысанын (мәтінді) нормалау сипатына аса көңіл бөлінуі қажет.

Сонымен, мәтінді ішкі бөліктеудің ең тиімді жолын табу лингвистикалық модельдеу ісінде аса маңызды деп саналады.

Енді қазақ мәтіндеріне қатысты жүргізілген кейбір тәжірибе нәтижелеріне тоқталайық [Джубанов, 1987].

Сөз таптарының тілде қолдану жиілігін сөз еткенде, барлық тілдерге де ортақ заңдылық – зат есім мен етістік сөздердің өнімді қолданылуы. Мысалы, М. Әуезовтің "Абай жолы" романы мәтінінде **зат есім** *35,36%*, ал **етістік сөздер** *31,44%*, яғни *100* пайыздың *70-не* жуығы тек осы екі сөз табына қатысты сөздер екен. Газет мәтінінің *45%* – **зат есім**, *22* пайызы – **етістік** (Ахабаев).

Сондықтан қазақ мәтіндеріндегі сөз үлестірілу заңдылығын зерттеу жұмыстарына алынатын таңдама мәтіндер әр сөз табына байланысты осындай қарапайым және тәуелсіз талаптарға жауап берулері керек. Егер аталған шарттар орындалды деп ұйғаруға болса, зат есім, етістік сөздердің (не басқа сөз таптарының) тәжірибелік үлестірілудің теориялық (нормальды не басқа) заңдылыққа үйлеседі деп ***"H_0"*** жорамалды жасауға болады. Бұл жорамалды "бағалау" әр түрлі ***N, n, k*** шамаларына қатысты $p(\chi^2 \geq \chi_q^2)$ ықтималдықтарды салыстыру негізінде іске асады.

Көркем әдебиет мәтіні, яғни "Абай жолы" романынан тәжірибе ретінде алынған ***N=50000*** сөзқолданыс:

а) $N_1=2500$, $N_2=5000$, $N_3=10000$, $N_4=20000$, $N_5=25000$, $N_6=40000$ және $N_7=50000$ көлемдері ретінде бөлек-бөлек қарастырылды;

ә) ішкі бөлік көлемі ***(n)***: $n_1=100$, $n_2=200$, $n_3=250$, $n_4=500$, $n_5=1000$ сөзқолданыстан тұратын көлемдері ретінде бөлек-бөлек қарастырылды;

б) барлық жағдайда $N=k \times n$ болатынын ескеріп, сериялар шамасы $k=N:n$ анықталды: $k_1=25$, $k_2=50$, $k_3=100$, $k_4=200$, $k_5=500$ сериялық сан мөлшері екені анықталды.

Осылайша нормалаудың нәтижесінде көркем әдебиет мәтіні бойынша **зат есім** сөздердің тәжірибелік жиіліктері оның теориялық жиіліктерімен сәйкес келудің "сенімділік дәрежесін" қанағаттандыру жағынан алғанда ***1-ші*** орында **"Нормаль үлестірілу"** заңы, ал ***2-ші*** орында **"Шарлье А үлестірілу"** заң түрі екені анықталды. Газет мәтіні бойынша, сериялар саны ***100-ден*** артқан жағдайда **зат есім сөздерің** теориялық заңдылықтарға өте сәйкес келу мүмкіндігі **"логарифмды-нормаль үлестірілу"** заңдылығы екені дәлелденді.

Көркем әдебиет және публицистика мәтіндеріндегі **етістік сөздердің** тәжірибелік үлестірілу деректерінің теориялық үлестірілу заңдарымен үйлесімділігін зерттей келе, олардың **зат есім сөздерге** сәйкес математикалық моделі (үлгісі) **нормаль заңдылығы** екені айқындалды. Ал егер, серия саны $k \leq 100$ болған жағдайда, етістік сөздердің үлестірілулері теориялық заңдылықтар ішінен **"Шарлье А үлестірілу"** типіне сәйкестігі айырықша байқалды.

Публицистикалық мәтіндердегі етістікке қатысты сөздердің тәжірибелік үлестірілуі параметрлерді нормалаудың мынандай түрінде: $\{N \geq 1000, n \geq 100, k \geq 100\}$, теориялық заңдылықтар ішінен **"нормаль үлестірілуін"** көбірек қанағаттандыратынын ықтималдықтың мәнімен байқауға болады.

Сонымен, қазақ тіліндегі **етістікке қатысты сөздердің** тәжірибелік жиіліктерінің мәтін бойындағы үлестірілуі параметрлерді аталған нормалауға сай, **"Нормаль заңымен"** үйлесімдік табады деп қорытындылауға болады.

Теориялық үлестірілу заңдарының **Пуассон** және **Шарлье В типі** қазақ тіліндегі **зат есім** мен **етістік** сөздердің тәжірибелік жиілік үлестірілу деректерімен үйлеспейтіндігі де байқалды.

Зат есім мен **етістік** сөздердің қатарында сын есім сөздердің де мәтіндегі үлестірілуі қарастырылды. Қысқаша айтқанда, олардың теориялық заңдылықтарымен сәйкестігі көркем әдебиет пен газет мәтіндерінде бірдей емес және нормалау түріне қарай, яғни ***N, n, k*** параметрлерін бөлшектеуге қатысты айырым табады. Дәлірек айтқанда, көркем әдебиет мәтіні үшін нормалау сипаты $\{N \geq 2500, n \geq 1000, k \geq 25\}$ болғанда, **"Пуассон үлестірілу"** мен **"Шарлье үлестірілу"** заңдылығының ***A*** мен ***B*** типтерімен бірдей дәрежеде үйлесетіні байқалды,

{N≥5000, n≥100, k≥50} ал газет тіліне қатысты нормалау әдісі ғана аталған заңдылықтарды қанағаттандырады екен.

Үйлесімдік дәрежесін оның **"артықшылық"** тұрғысынан қарастырсақ, әрине, **сын есім** сөздерге қатысты **Пуассонның теориялық үлестірілу заңы** бірінші орынға ие. Солай бола тұра, нормалаудың ***{N≥2000, n≥100, k≥200}*** мәндерінен бастап, қазақ тілінің **сын есім сөздері "нормаль заңдылығымен"** де жақсы үйлесімдік табатынын байқаймыз.

Сөз таптарының тәжірибелік жиіліктерінің теориялық үлестірілу заңдылықтарымен үйлесімдігі, көбіне, сериялар санына қатысты екені байқалады. **Үстеу** сөздерге қатысты ең "қолайлы" деп **Пуассон үлестірілу** заңын атауға болады. Ал сериялар саны ***k>500*** болғанда үлестірілу **"нормаль заңымен"** үйлеседі.

Сонымен, жоғарыда баяндаған жайттар мәтінді мөлшерлеу (нормалау) әдістерінің қазақ тіліндегі сөз таптарының тәжірибелік үлестірілулерінің белгілі бір теориялық үлестірілу заңдылығына сәйкес келу не келмеу мүмкіндіктерін айқындауды көздейді. Бұл зерттеудің нәтижелері қазақ тіліндегі құбылыстарды модельдеу мәселелерінде аса маңызды рөл атқаратыны сөзсіз.

Алтыншы тарау
ҚОЛДАНБАЛЫ ТЕРМИНТАНУ

6.1 Қолданбалы терминтану саласының негізгі ерекшеліктері

Қолданбалы лингвистиканың мәселелерін зерттеуде оның терминтану жағына айрықша көңіл бөлінуі қажет. Терминтану қазіргі заман сұранысына сай ғылыми пән және оның зерттеу нысаны – терминдер мен терминологиялық жүйелер.

Бұл ғылым саласы XX ғасырдың ортасында тіл білімінің, логиканың, информатиканың, психологияның және басқа да ғылымдардың түйісу жерінен пайда болды. Қай тілді алмасаңыз да терминдер – сөз немесе сөз тіркестер түрінде кездеседі де осы тәріздес басқа сөздердің заңдылығымен бірдей қолданыс табады. Сондықтан терминтану пәнінің әдіс-тәсілдері тіл білімінің ұстаным ететін зерттеу жолдарымен ұқсас болып келеді. Мысалы, терминтану пәні де лингвистикадағыдай *теориялық* және *қолданбалы* деп аталатын екі бөлімнен тұрады.

Терминтанудың қолданбалы жағын сөз еткенде, ең алдымен, оның терминтану саласындағы мәртебесін анықтап алу қажет. Тіл біліміндегі қолданбалы лингвистиканың өз теориясы болатындығы сияқты қолданбалы терминтанудың да теориясы және оның практикалық терминологиялық қызметінің бірқатар бағыттары бар. Бұл бағыттардың өздеріне тән терминологиялық мектебі бар десек те болады.

Сондықтан терминологиялық бағытты таңдау оның қолдану қағидаттарына, әдістеріне, көлеміне, қызмет ету деңгейіне, халық шаруашылығындағы нәтижелеріне, экономикалық сипатына және т.б. байланысты болады. Осыған орай, терминтану мектептерінің практикалық қызметінің мынадай түрлерін атауға болады [Лейчик, 1996, 276-286-бб.]:

- лексикографиялық терминология қызметі;
- термин мен терминдік жүйелерді жетілдіру (реттеу, тұрақтандыру, редакциялау);
- ғылыми-техникалық және басқа да маманданған мәтіндер аясындағы терминдерді аудару;
- терминологиялық деректер банкін құру (ТДБ);
- терминологиялық ұйымдар мен орталықтардың методологиялық (ғылыми) қызметін ұйымдастыру және т.б.

Терминологиялық қызметтің нәтижелерін пайдалану аясы терминтану саласына күн сайын жаңа мәселелер қоюда. Мәселен, аса көлемді арнайы ақпаратты сақтау мен өңдеуді автоматтандыру қажеттігі мен оны іске асырудың ақиқаттық мүмкіндігі "Терминологиялық деректер базасын" (ТДБ) өмірге әкелді. Осыған қатысты мынадай қызықты және әлі де тың мәселені де айта кетуге болады. Ол – терминологиялық стандарттар мен оларды ұғыну мүмкіндіктері жайлы ұсыныстар төңірегіндегі пікірталастардың өз бетінше стихиялы түрде пайда болуы.

Сонымен бірге жаңа техника мен технологияны өмірге келтіруге қатысты жұмыстардың туындатып жатқан жаңа нысандарды, құбылыстарды, ұғымдарды

анықтайтын атауларды ұсыну мәселесі. Сондықтан атау беру (номинация) проблемасы – практикалық терминологиялық қызметтің өзінше бір бөлек, маңызды мәселесі. Себебі, жаңа терминдер атауларын жасау үшін семантикалық, сөзжасамдық, синтаксистік, аударма (калькілеу, транслитерациялау) заңдылықтарына сүйену қажет.

Бір жағынан алғанда, бір нәрсені атау (номинация) мәселесі тек терминдер аясына ғана тән емес, ол – табиғи тіліміздегі "ат берудің" барлық жағдаятына ортақ іс.

Екіншіден, атау ұсынуда терминологиялық, лингвистикалық және логикалық әдіс-тәсілдер қабаттасып, араласып жатып, пікірталас жағы терминнен гөрі, ұғымдар жағына ауысып кетеді. Яғни бұл мәселенің таза терминологиялық қызметінен сырт жатқанын байқаймыз.

Үшіншіден, ең маңызды жағы, аталған қызметтер кезінде саналы не санасыз түрде терминжасам заңдылықтарын білу қажет болады, сондықтан, ең дұрысы, терминологиялық атау мәселесі (номинация) – практикалық үлесі бар теориялық қызмет. Бұл жердегі практикалық үлес деп отырғанымыз – терминжасау, сөйлеу тілі мен терминдік жүйе құрамына енгізу.

Практикалық терминологиялық қызмет өзара тәуелді екі түрлі аспектіге ие: оптимизациялық және нормативтік. Бір жағынан алғанда, ондай қызмет терминдер мен олардың мағынасын дәл анықтайтын және тұтынушыны қанағаттандыратын тиімді вариантты таңдап алуға бағытталған. Екіншіден, таңдалып алынған тиімді терминдер мен олардың анықтамалары нормативті құжаттар мен бұйрық-жарлықтар қағаздары арқылы бекітіліп, баянды етіліп отырады.

Оларға жататындар: стандарттар, терминологиялық кепілдемелер, нормативті сөздіктер, редакторлық нұсқаулар мен ережелер және т.б. Бұл құжаттар мен бұйрық-жарлықтар терминологиядан тыс жатқан лексика аясында қолданыс табатын нормаға қарағанда әсерлілеу, күштілеу деуге болады.

Міне, бұдан да терминтану ғылымының лингвистикадан айырмашылығын байқауға болады. Себебі, терминтану ілімі лингвистикаға жат нормативті құжаттармен әрекет етеді және бұл құжаттар заңдылық мәртебеге ие, ал мұндағы терминдер өндірістік нормамен бірдей дәрежеде қолданылады.

Бұл қысқаша баяндалған мәселелерден қолданбалы терминтану мен оның бағыттарының қолдану аясындағы қайсыбір жалпы ерекшеліктерін аңғаруға болады.

6.2 Терминтану пәнінің негізгі бағыттары

Төменде аталатын бағыттардың бір-бірімен өзара қатыстылығын, яғни олардың барлығының бірдей бір-біріне ауысып тұратындығын және бірін-бірі анықтайтындығын есте ұстау қажет (сөздіктен стандартқа және т.б.).

6.2.1 Ғылыми-техникалық лексикография

Ғылыми-техникалық лексикографияның зерттеу нысаны – арнайы сөздіктерді (терминологиялық) құрастырудың теориясы мен практикасының мәселелері. Арнайы сөздіктер адамға немесе электронды машиналарға бағыттальп түзіледі. Бірінші аталғанға біртілдік және аударма терминологиялық сөздіктерді, идентификаторлық сөздіктерді, терминдердің жиілік сөздіктерін, рубрикаторларды және

классификаторларды жатқызуға болады. Ал екіншіге – әр типті ақпараттық жүйелер үшін түзілген ақпараттық-іздеу тезаурустары, мәтінді автоматты өңдеудің әр түрлі жүйелеріне арналған машиналық сөздіктердің түрлері жатады. Түсіндірме сөздіктерге қарағанда, терминологиялық сөздіктер белгілі білім және техника саласындағы мамандарға арналып құрастырылады. Атқаратын міндетіне қарай терминологиялық сөздіктер, бір жағынан, жоғары білімді мамандарға, ғылыми әдебиеттерді аударатын кәсіби аудармашыларға арналған. Мұндай сөздіктер: жоғары квалификациялы мамандарға қажетті деген сөзді тауып алу мүмкіндігімен бірге, оның белгілеп тұрған ұғымын анықтауға, неше түрлі мақалаларда немесе кітаптарда кездесетін әрқилы синонимдік қатарларды ажыратуға мүмкіндік жасауы керек. Ол үшін үлкен идеографиялық (тақырыптық) ұстанымда түзілген біртілді және көптілді түсіндірмелік-аударма терминологиялық сөздіктер қажет. Екінші жағынан, студенттерге, орта өндіріс пен ауыл шаруашылығындағы кәсіби мамандарға, теле-радио мен баспа қызметкерлеріне арналған біртілді және аударма сөздіктерінің де арнайы түзілгені жөн. Мектеп оқушыларына арналып түзілген терминологиялық сөздіктің орны айырықша. Мұндағы негізгі ұстаным – жас ерекшелік. Оқырманның жасына қарай сөздік материалдарының берілу тәсілдері де ерекшеленеді. Мысалы, оқырманды қызықтыру элементтерінің болуы, суреттердің орын алуы, неше түрлі сызбалармен безендірілуі, нақты есептерді ұсыну жолдары және т.б.

Соңғы кезде ғылыми-көпшілікке арналған терминологиялық сөздіктер де жарық көруде. Бұл сөздіктерде әр ғылым саласының терминдік жүйесі көрініс табады. Әдетте мұндай сөздіктер әліпби тәртібімен емес, жүйелік ұстанымға негізделіп құрастырылады.

Мақсаты жағынан мынадай терминологиялық сөздіктерді атап өтуге болады: "узус" сөздіктері (тіркеу, түгелдеу сөздіктері), регламенттеу, оқу жұмысы, хабарлау, жүйелеу мақсатындағы сөздіктер.

Ғылыми-техникалық лексикография негіздерін толығырақ меңгеру үшін А.С.Гердтың "Основы научно-технической лексикографии: как работать над терминологическим словарем" атты еңбегіне сілтеме жасағанды жөн көреміз (Герд, 1986.73-б.).

6.2.2 Термин мен терминдер жүйесін бірізге келтіру

Бұл бағыт қолданбалы терминтану саласының негіздерінің бірі болып саналады. Бір ізге салу жұмысының жемісі ретінде регламенттеу сөздіктерін, нормативті және нормативті емес сөздіктерін атауға болады. Нормаға қатысты сөздіктерде терминдердің әр категорияларына стандарт көрсетіледі. Сонымен бірге бұл реттегі орындалған жұмыстарға ұсынылған терминдердің жинағын да жатқызуға болады.

XX ғасырдың 90-жылдарының басында, әлем бойынша, терминдер мен олардың анықтамаларына стандарттаудың саны 20 мыңнан асты. Олардың негізгі бөлігі өндірістері дамыған елдерде зерттеп дайындалған – ұлттық стандарттар. Бұл айтылғаннан басқа: халықаралық стандарттар, аймақтық стандарттар, фирмалар мен халықаралық ұйымдардың стандарттары да қолданыс табуда. Терминологиялық стандарттарға қойылатын талаптардың көпшілігі әртүрлі құжаттарды жазуда міндетті түрде орындалуы қажет. Ал кейбір талаптардың арнайы ұсыныс арқылы іске асатындары да бар.

Реттелген терминдер ұсынылған (бекітілген) терминдер жинағы түрінде жарық көреді. Ресейде ондай терминдер Ғылым академиясы жанындағы Ғылыми-техникалық терминологиялық комитетінің бекіту жарлығымен жұртшылыққа ұсынылады. Ал Қазақстанда мұндай жұмыстарды Мемлекеттік терминологиялық комитеті атқарады.

Бұл бағытқа, яғни терминдер мен олардың жүйесін бірізге келтіру мен оларды арнайы құжаттар арқылы бекітуге қарсы пікірлер де жоқ емес. Бірақ машина мен механизмдердің бөлшектері мен тораптарын стандарттау қаншалықты қажет болса, терминдерді де бірізге келтіру ғылым мен техниканың алға басуына соншалықты қажет деуге болады. Егер терминдердің атаулары арасындағы синонимдік сипат жойылатын болса, онда ол ғылыми қызметтің "тоқырап қалуына" әкеліп соғады деген күдік те туындауы мүмкін. Ал бірізге түскен терминдерге қатысты нормативті құжаттардың уақыт пен кеңістік бойынша және оның қызметіне қарай шек қойылу жағдаяты жоғарыда аталған күдікті жоққа шығара алады.

Терминдерді бір ізге салудың тағы бір оң қызметі – жүйелілікке келтіру мүмкіндігі. Стандарттар мен ұсынылған терминдердің жинағында, сөздіктерге қарағанда аталған қызмет басым байқалады. Стандарттау мен ұсынылған терминдерді жинақтауды дұрыс жолға салу нәтижесінде әрбір терминдік жүйедегі орын бір ғана лексикалық бірлікпен толтырылып, оған сәйкестікке бір ғана анықтама беріледі (сөздіктің оң жағындағы мақаласы). Әрине, бұл әрекет, белгілі білім аясына жататын барлық таңбалау құралын жүйеге келтіреді және мұндай білім аяларына шектеулер қойып, олардың өзара байланыстығы анықталады. Терминдерді бір ізге салуда мұндай шараның да маңызы зор.

Соңғы жылдары терминдер мен терминдік жүйені бірізге салудың үшінші түрі орын ала бастады – ол, ұлттық және халықаралық деңгейдегі үйлестіруді іске асыру. Әрине, бұл әрекет бойынша әртүрлі тілдердегі және жақын білімдер аясындағы терминдерді бүтіндей теңестіру мақсаты қойылмайды. Дегенмен, терминжасау мамандары мен әр ғылым салаларының өкілдері әр дәрежеде терминдердің өзара үйлесімділігін іздестіруге және оның шегін анықтауға атсалысады. Терминдер қолданыс табатын тілдердің өзара ерекшеліктері ескеріледі – сөзжасам тәсілі, фонетикалық айырмашылығын, сөздердің тіркесім заңдылықтарын және т.б. Сонымен бірге әрбір терминдік жүйе белгілі теорияға, тұжырымдамаға, көзқарастар жүйесіне тәуелді екені де ескеріледі. Мысалы, физика саласында, ең аз дегенде, екі терминдік жүйе орын алады, оның біріншісінің негізінде – Ньютонның классикалық механика теориясы жатыр, ал екіншісінің негізінде – Эйнштейннің салыстырмалы теориясы жатыр. Сондықтан бір ғана термин, мысалы, "масса" термині аталған екі жүйеде бөлек-бөлек мағынада қолданылады.

Осы айтылғандардан барып, терминдік жүйелердің үйлесімділігін қарастыруда екі топқа енетін факторлар ескерілуі қажет: терминдер қолданылған тілдердің ерекшелігіне байланысты, яғни лингвистикалық және пәндік сала мен оны сипаттайтын теорияға қатысты – экстралингвистикалық фактор.

Терминдік жүйелердің халықаралық үйлесімділігінің экстралингвистикалық алғышартына жататындар: білімнің интеграциясы, ғылым мен технологияны интернационалдау, ғылым мен технологияның теориясы мен әдістемелік базасының бірлігі. Терминдік жүйелердің халықаралық үйлесімділігіне мүмкіндік туғызатын лингвистикалық факторларға жататындар: арнайы мақсатпен бір-біріне жақын жатқан табиғи тілдер үшін ортақ тіл туындату, арнайы мақсаттағы тілдердің лексикалық

бірлігінің функциясы ретінде орын алатын терминдерді интернационализациялаудың жедел әрекеттері.

Терминдік жүйелердің халықаралық үйлесімділігінің негізгі екі қағидаты бар:

а) егер екі не одан да көп мемлекеттерде білімнің белгілі бір саласы немесе сол саладағы зерттеулер бірдей дәрежеде дамыған болса, аталған үйлесімділікті қабылдаудың мүмкіндігі бар және ол мақсатқа сай келеді;

ә) егер бұл мемлекеттерде белгілі бір білім саласының негізінде бірдей не жақын жатқан ұғымдар жүйесінің ұқсас (бірдей) немесе жақын жатқан теориясы болса, аталған үйлесілімділікті қабылдауға мүмкіндік бар.

Аталған ұстанымдар, белгілі бір дәрежеде, іске асып жатса, терминдік жүйелердің мынадай үш дәрежедегі үйлесімділігіне қол жеткізуге болады:

1) теориялардың ішінара сәйкес келуі мен тілдік құралдардың айырым табу жағдайындағы **ұғымдар жүйесінің ортақтығы**. Бұл жағдайда, үйлесімділік тек ұғымдар жүйесі мен оларды белгілейтін терминдерді өзара байланыста ұстайтын тұрпатта түсінік табады;

2) теориялар мен тілдік құралдардың айырым табу жағдайының толық сәйкес келуіндегі **ұғымдар жүйесінің ортақтығы**. Бұл жағдайда, үйлесімділік тек ұғымдар жүйесінің бірдей келуі мен оларды белгілейтін терминдердің құрылымдық-семантикалық жақындығы түрінде түсіндіріледі;

3) теориялар мен тілдік құралдардың толық сәйкес келу жағдайындағы **ұғымдар жүйесінің ортақтығы**. Бұл жағдай, үйлесімділік ұғымдар жүйесі мен терминдерді интернационализациялаудағы мазмұнның көлемі мен өзара байланысының бірдейлік сипатымен көрініс табады.

Терминдік жүйелердің халықаралық үйлесімділігінің қолданбалы нәтижелеріне жататындар: халықаралық терминологиялық сөздіктер мен стандарттар және көптілді ақпараттық-іздеу тезаурустары.

6.2.3 Ғылыми-техникалық және басқа да арнайы терминдерді аудару мәселесі

Арнайы мәтіндерді аударуда, оларда қолданылған терминдерді барабар аудару ең маңызды іс болып саналады.

Аудармаға арналған көптеген ғылыми конференциялар мен семинарларда терминдерді аудару тәсілін таңдау проблемасы төңірегінде өткен пікірталастардың маңызы зор. Қазіргі кезде ғалымдардың осындай әртүрлі тәсілдерді қолдану жайлы ұсыныстары зерттелуде (Лейчик, Шелов, 1989-1990). Оларды қысқаша былай тұжырымдауға болады:

а) Ең алдымен аударма тілі (АТ) және түпнұсқа тілі (ТТ) деген атаулардың мағынасын ажырата білуіміз керек. Егер "аударма тілінен" (АТ) белгілі бір тәсілге сүйеніп, "түпнұсқа тіліне" (ТТ) сәйкес келетін қолайлы балама табылып жатса, сол тәсілді оңтайлы тәсіл деп атауға болады. Мұндай тәсілді қолдану үшін тіл иелеріне қатысты мемлекеттердің қоғамдық даму деңгейлері бірдей немесе өзінің тарихи кезеңдерінің бірінде одан өтіп кеткен болуы керек. Мысалы, екі тілдегі балама-терминдердің ерекшелігі мынада. Әдетте, әр тілдегі балама-терминдер екі тілде де бұрын қолданылған болса, аударма жасау кезінде жаңа терминдер мен неологизмдер сияқты оларды қайта жасамай, баламаларын қайтадан іздеп-табуға әрекет етуіміз керек.

ә) Аударма тіліндегі (АТ) жаңа терминді жасау үшін осы тілде бар сөз не сөз тіркеске түпнұсқа тілі (ТТ) әсерінің арқасында жаңа мағына берілуі қажет. Мысалы, орыс тіліндегі ботаникалық термин *дерево* жаңа мағынаға ие болып, жаңа ғылыми салалардың – классификациялау теориясы, басқару теориясы, ақпарат теориясы терминдеріне айналып отыр. Оны мынадай терминдік сөз тіркестерден байқауға болады: *дерево целей, поисковое дерево* (ағылшын тілінің *tree* сөзінің әсерінен) және т.б. Орыс тілінің саяси термині *"разрядка"* бұрын физика саласында жиі қолданылып жүрген терминнің жаңа мағынасы деуге болады. Бұл сөздің жаңа терминдік мағынаға ие болуы француз тілінің саяси мағынадағы *detente* сөзіне қатысты екен. Себебі *detendre* – "*ослаблять, разряжать*" етістігінен жасалғаны анық.

б) Егер аударуға тиісті лексикалық бірлік екі тілде де сәйкес келіп жатса, оларды ***семантикалық калька*** деп атайды. Семантикалық кальканы пайдаланғанда түпнұсқа тіліндегі (ТТ) терминнің құрылымы түпнұсқа тілінің нормасына сәйкес келеді, ал аударма тілінің (АТ) жасалатын термин құрылымы – аударма тілінің (АТ) нормасына сәйкес келеді. Техникалық ғылымдар саласында құрылым жағынан күрделі болып келетін терминдердің әр элементін бөлек-бөлек аудару (калькілеу) тәсілі кең қолданылады. Мысалы, ағылшынша: *Swithing diagram – коммутационная схема, motor selector – моторный искатель* (электроника бойынша ұсынылған мемлекеттік стандарттан алынады).

в) Егер аударма кезінде лексикалық бірлік сол құрылымы қалпында басқа тілдегідей қайталанып пайдаланылса, ол құрылымдық калька немесе кальканың өзі деп саналады. Кейбір басқа тілден ауысып алынған құрылымдық модельдер калькілеу нәтижесінде аударма тілінде (АТ) тұрақталып қалады, ал басқалары жат қалпында қалып, жекеленген құрылымдық калька ретінде қолданыс табады. Міне, калькілеу тәсілінің ұтатын да, ұтылатын да жері осында.

г) Егер аударма кезінде түпнұсқа тіліндегі (ТТ) терминдердің әрі семантикасы, әрі құрылымы, әрі тұрпаты (дыбыстық құрамы мен жазылуы) толығымен аударма тіліне (АТ) өзгеріссіз ауысса, ол кірме терминдер болып саналады. Осындай жағдайда, кірме терминдерді екі тілдің тікелей түйісуінен болатын және интернационалдық құбылыстардан нақты ажырата білген жөн. Ең бірінші, кірме терминдерді классикалық жасалу тәртібіне негізделген еуропа мәдениетінің ерекшелігімен анықталатын грек-латындық элементтер арқылы құрылған тілдік бірліктермен шатастырмауымыз керек.

Аударма тұсында нағыз кірме сөз-терминге де көзқарас бірдей емес. Егер термин аударма тіліне өзі анықтайтын немесе түпнұсқа тілінде белгілейтін жаңа нысанның ұғымымен бірге "келсе", ондай кірме терминді қабылдауға болады. Мысалы, ағылшын тілінің американдық қолданылуындағы *джинсы* (ағыл. jean – комбинезон, джинсилық мата; jeans – джинсы) термині орыс тіліне де еніп отыр. Сол сияқты орыс тіліндегі *самовар* сөзі француз тіліне ауысты және т.б. Мұндай "эквивалентсіз лексика" деп аталатын кірме сөздер аударма кезінде, қисындылық және әдістемелік тұрғыдан қарастырғанда өзін-өзі ақтайды. Бұл айтылғандар адамның рухани өміріне де бірдей қатысы бар: француз тіліндегі *patriots* орыс тілінде – *патриот;* ағылшын тіліндегі *soviet* орыс тіліндегі *совет* сөзінен ауысқан. Бірақ, егер түпнұсқа тілі (ТТ) мен аударма тілі (АТ) қолданылатын мемлекеттерде ғылым мен техника деңгейі бірдей болып келетін жағдайда, аудармашы мен редактор **түпнұсқа тілінен аударма тіліне** терминнің баламасын немесе аударма тілі элементтері

ішінен жаңа термин табуға ынталануы қажет. Мысалы, ағылшын тіліндегі *know-how* терминін *ноу-хау* деп алғаннан гөрі, оны орыс тіліне аударып – *секреты производства* немесе қазақ тіліне аударғанда – *өндіріс құпиясы* деп алғаны жөн болар еді.

д) Көптеген жағдайда термин сөз – сипаттау құрылымымен бірге аударылғаны дұрыс деп саналады. Бұл аудару әдісі, ең алдымен, нақты мемлекеттің ақиқат өмірін суреттейтін эквивалентсіз терминдер үшін қолданылады. Мысалы, Кеңес заманы кезіндегі *ударник* термині Уэбстрдің студенттерге арналған сөздігіндегі суреттеу конструкциясы ең дұрыс шешім деп танылды: *a workes who had excelled in voluntary increase of production – "рабочий, который на добровольных началах добился высоких показателей в производстве"* (қазақшасы – *"өндірісте өз еркімен жоғары көрсеткішке қолы жеткен жұмысшы"*.

6.2.4 Терминологиялық редакциялау

Мәтіндерді әдеби және техникалық редакциялаумен қатар терминологиялық редакциялау – мәтінді ғылыми редакциялаудың бір түрі. Терминологиялық редакциялауды қарастыру үшін оның мақсаты мен міндетін, негізін және әдістерін сипаттап алу қажет.

Редакциялаудың басқа да түрлері сияқты оның мақсаты – мәтіннен танымдық, білім беретін немесе тәрбие алатын, яғни барынша толық түрдегі күшті әсер алатын зор нәтижеге жету.

Терминологиялық редакциялау барысында алға қойған мақсатқа жету үшін ерекше материал ретіндегі терминдермен әрекет етуге тура келеді. Ал терминдер, жоғарыда айтылғандай, белгілі тілдің табиғатына, оның терминдік жүйеден алатын орнына көп байланысты. Сондықтан терминологиялық редакциялаудың міндеті – ол, терминдер мен сөздерді және олардың орнына жүретін тілдік бірліктерді барынша орынды пайдалануға мүмкіндік туғызу. Терминдерді таңдауда, оның нақты мәтіндегі ақпараттық рөлін және сол ғылым не техника саласындағы басқа терминдермен салыстырғандағы орнын ескеру қажет болады.

Терминологиялық редакциялаудың қағидаттары көптеген жағдайда шығарманың стилі мен жанрына байланысты болып келеді. Мысалы, ғылыми мәтіндерде (ғылыми мақала, монография) арнайы терминдер көптеп кездеседі де, бірақ маман-оқырман оларды жақсы білуі керек деген көзқараспен терминдерге арнайы түсініктеме берілмейді. Ал ғылыми-көпшілікке арналған мақалалар мен газет бетіндегі очерктерде терминдерге тиісті сөз орамдарымен сипаттамалар беріліп отырады. Себебі көпшілік оқырман мәтінде баяндалған ғылыми проблемамен, саяси немесе экономикалық оқиғалармен бірінші рет танысып отыруы да мүмкін.

Терминологиялық редакциялаудың әдістеріне келетін болсақ, олардың саны көп жағдайда екіден аспайды. Біріншісі бойынша, редактор мәтінде қолданыс тапқан терминдерді тиісті анықтамалармен немесе бекітілген нормативті құжаттармен (стандарттар, ұсынылған терминдер жинағы), не болмаса осы ғылым мен техника аясында кең қолданылатын терминдермен сәйкестікке қояды. Ал, екінші әдіс бойынша, орын алатын әрекет мынаған тіреледі: егер редакцияланудан өтіп жатқан мәтін ішінде жаңа терминдер кездессе, автор олардың не себепті орын алғанын түсіндірумен қатар, әрбіреуіне анықтамалар беріп отырады.

6.2.5 Терминологиялық деректер базасын құру

Бұл бағыт – терминологиялық қызмет жасаудағы ең бір заман сұранысына сай, жаңа бағыт болып саналады. Мұндай бағытқа бастама болған терминологиялық көртішкілерді жинақтау мен олардағы ғылыми, техникалық, экономикалық және т.б. пәндерге қатысты көртішкілерді қажеттікке сай тауып алу жұмысы. Сонымен бірге, ондай көртішкі көлемінің үнемі өсіп отыру сипатына қарай, қажетті ақпаратты тауып алу мен оларды өңдеу жұмыстарының барынша қиындауы да негізгі себептердің бірі деуге болды. Аталған жұмыстарды жеңілдету үшін ондай іс-әрекеттерді компьютерге жүктеу қажеттігі туындады. Себебі қазіргі кезде аса көлемді деректерді сақтау мен оларды өңдеу және қажетті деген ақпаратты тез арада іздеп-табу жұмыстарын тек компьютерге жүктеу арқылы іске асыруға болатыны белгілі. Бұл бағыт бойынша атқарылатын жұмыстың маңыздылығы – мамандар мен ғалымдарды нақты және нормативті терминологиялық құжаттар арқылы бекітілген деректермен дер уақытында қамтамасыз ету. Міне, осы аталған жайттар Ресей мен алыс-жақын шет елдерде көптеген компьютерлік терминологиялық деректер базасын (ТДБ) жасауды қажет етті. Әрбір ТДБ қосымша ақпараттармен қамтамасыз етіледі: терминнің мағынасына қатысты олардың анықтамасы, басқа тілдердегі баламалары, терминнің қысқартылған пішіні, оның мүмкін болатын және мүмкін болмайтын синонимдері, терминдердің қандай құжат негізінде бекітілуі (мысалы, ұлттық не халықаралық стандартпен) және т.б.

ТДБ құрудың мақсатына қарай, оларды екі топқа бөлуге болады: ғылыми-техникалық әдебиет пен құжаттарды аударуға бейімделген бағытта және терминологияның стандарттануы мен ұсынылуы жайлы ақпаратпен қамтамасыз ету бағытында.

Қазіргі кезде бірнеше ондаған ірі ТДБ іске қосылып отыр. Ресейде стандарттаумен айналысатын мекемелерде "стандартталған терминдер банкі" деп аталатын деректер базасы жұмыс істейді (ВНИИКИ институтында), медициналық терминдер банкілері және т.б. Әлем бойынша, ең ірі ТДБ – германиялық "Сименс" фирмасы болып саналады. Бұл терминологиялық базада 8 тілді қамтитын (орыс тілі де) 2,5 млн жуық терминологиялық деректер жазылған. Сол сияқты, мұндай компьютерлік базалар Францияның стандарттау мекемесінде, Канадада екі банк іске қосылған. Люксембург қаласында Еуропа одақтастығы комиссиясының терминологиялық бюросында бірнеше тілдердегі автоматтанған терминологиялық сөздік түрінде түзілген халықаралық ірі ТДБ құрылған. Оның қысқарған аты – "Евродикаутом".

Жоғарыда сөз болған терминологиялық деректер базасының (ТДБ) теориялық та және практикалық та болашағы зор деп саналады. Ол автоматтанған "қойма" болудан әрі қарай дамып, аса маңызды ғылыми-зерттеу құралына айналып отыр.

6.2.6 Ұйымдастыру және әдістемелік шаралар

Тақырыпта аталған шаралар түрлері мыналар: терминологиялық деректер базасын (ТДБ) құруда терминологиялық стандарттар мен сөздіктерді түзуге жетекшілік ету, терминология мәселесі жайлы ғылыми конференцияларды, семинарларды және т.б. ғылыми кездесулерді өткізу, әртүрлі мамандар

категорияларын (қолданбалы лингвистика аясындағы аудармашылар, стандарттаушылар, ғылыми қызметкерлер) терминтану негіздеріне оқыту.

Осы аталғандар және басқа да ұйымдастыру-методикалық қызмет түрлерін ұлттық және халықаралық ұйымдар атқарады.

Енді қай жерлерде қандай ұйымдар терминологиялық қызметпен айналысатыны жайлы қысқаша тоқталайық.

Ресей астанасы Мәскеуде: терминологияны стандарттау проблемасына жауап беретін және "Научно-техническая терминология" атты ғылыми журналын шығарып тұратын "Стандарттау мен сапа жөніндегі комплексті ақпараттардың ғылыми-зерттеу институты" (ВНИИКИ), жаратылыстану мен техника ғылымдары аясындағы терминологияны реттеуге жетекшілік ететін РҒА-ның "Ғылыми-техникалық терминология комитеті"; арнайы сөздіктер шығаратын "Русский язык" баспасы және т.б. баспалар; медициналық терминдер банкісін құрумен айналысатын және медициналық терминологиялық сөздіктер мен анықтағыштар шығарып тұратын "Медицинская энциклопедия" ғылыми-өндірістік бірлестігі. Сол сияқты Мәскеу қаласындағы қоғамдық ұйымдарды, соның ішінде ғылыми-техникалық терминологияның методологиялық проблемалары бойынша тұрақты әрекет ететін семинарды, медициналық терминология проблемалары және т.б. бойынша жұмыс істейтін семинарларды атап өтуге болады.

Санкт-Петербург университеті мен Омбы қаласының терминологиялық орталығы да ТДБ мен терминологиялық сөздіктерін түзуде көптеген жұмыстарды атқарып жатыр. Нижгород университетінің "Термин и слово" атты ғылыми еңбектердің жинағын шығарып тұратыны жайлы да айта кетуге болады.

Тәуелсіз мемлекеттердің, мысалы, Украина, Белорусия, Латвия, Қазақстанның ғылыми орталықтарында да терминтанудың қолданбалы мәселелері бойынша үлкен жұмыстардың жүргізіліп жатқаны мәлім.

Халықаралық ұйымдар ішінен Вена қаласында орналасқан терминологиялық ақпарат орталығын (Инфотерм), Варшавада орналасқан және "Neotern" журналын жарыққа шығаратын терминологиялық неологизмдерді жетілдіру Халықаралық орталығын (МОУТН) атауға болады. Сонымен қатар халықаралық стандарттарды зерттеу жұмысымен айналысатын Халықаралық стандарттау ұйымын (ИСО), Халықаралық электротехникалық комиссияны (МЭК) және т.б. қызмет көрсетеді.

Бұл айтылғандардан мынадай қорытынды жасауға болады. Қазіргі кезде бүкіл әлем бойынша теориялық және практикалық тұрғыда терминологиялық қызметке барынша көп көңіл бөлінуде. Мұның себебі, ғалымдар мен өндіріс мамандарының рационалды және дәл сипаттағы терминологиясынсыз ғылым мен техникада, коммерция мен экономикада айтарлықтай жетістікке жету мүмкін емес.

Жетінші тарау

ҚАЗАҚ МӘТІНІНІҢ АБЗАЦТЫҚ ҚҰРЫЛЫМЫНЫҢ ЫҚТИМАЛДЫ-СТАТИСТИКАЛЫҚ МОДЕЛІ

7.1 Қазақ мәтінінің абзац деңгейіндегі функционалды-мағыналық құрылымының статистикасы

Егеменді елімізде қазақ тілінің мемлекеттік тіл ретінде дамуына зор мән берілуі тілші ғалымдардың өз ісіне жаңаша көзқараспен қарауын қажет етеді. Оған ең негізгі себеп – тілді компьютерлік жаңа технологияның жетістіктерін пайдалана отырып зерттеудің, оның нәтижелерін дұрыс тұтыну. Міне, осыған байланысты қазақ мәтіндерінің өзіне тән жасалу немесе туындау моделін құру – қазіргі заманымызға сай, кезек күттірмейтін мақсаттардың бірі.

Қазақ тілінде көптеген іс қағаздарының, неше түрлі хабарламалардың, ғылыми-техникалық әдебиеттер мен мектеп және жоғары оқу орындарына арналған оқулықтар мен оқу құралдарының жарық көре бастағаны бәрімізге мәлім. Олай болса, қазақ тіліндегі мәтіндерді компьютер көмегімен автоматты түрде өңдеу мен қажеттілікке сай пайдалану, тіл зерттеу тәжірибесіндегі жаңа сала – мәтін теориясымен тығыз шұғылдануды керек етеді. Жедел дами бастаған мәтін теориясы көптеген күрделі проблемаларды кеңінен қарастыруға жол ашты. Бұл тұрғыдан қазақ тілінде көрініс тапқан мәтіннің негізгі бірліктерінің мәнін ашу және олардың өзара тәуелділігінің сыр-сипатын, тақырып рөлін ашу мәселелері тілші ғалымдар үшін аса маңызды мәселе болып тұр. Дәлірек айтқанда, оқырманның көлемді мәтіндердің мағынасын аз мезгілде дұрыс ұғына білуі үшін және оқулықтар мен оқу құралдарын жазу кезінде мәтін құрылымының заңдылығын білген жөн. Міне, сондықтан да қазақ мәтінінің ішкі заңдылықтарын егжей-тегжейіне дейін қарастырып, зерттеу – заман талабы деуге болады. Бірақ бұндай зерттеулердің өзіндік қиыншылықтары жоқ емес.

Осыдан біраз жылдар бұрын мәтін ұғымы ешбір анықтамасыз-ақ қабылданып, ал тіпті берілген анықтамалардың өздері де бірізді болмай жүргені, бұл істің оңайға түспейтіндігіне дәлел болар деп ойлаймыз. Ондай анықтамалардың кейбіреулеріне мысал келтірейік: "мәтін дегеніміз ұзын сөйлем"; "мәтін – ол бір-бірімен мағыналық қатынаста болатын, кемінде екі сөйлемнен тұратын сөйлемдер тізбегі"; "мәтін – тұтас бір мағыналы белгілер құрылымы"; "мәтін – семантикалық жағынан ұйымдасқан белгілер жүйесі" және т.б. Ал, логика-лингвистикалық тұрғыдан құрылған анықтама бойынша, мәтіннің басты белгісі ретінде сызықтық тізбек құрайтын сөйлемдер алынатындығы және олардың сол жақтық, оң жақтық шекаралары болатындығы, салыстырмалы түрдегі толықтық пен тәуелділік шарттарының орындалуы тәрізді ерекшеліктер басшылыққа алынады.

Осындай анықтамалар негізінде біз өз тарапымыздан мәтіннің мынадай негізгі қасиеттерін атап кеткенді жөн көріп отырмыз. Ол – мәтіннің бір мазмұнды болуы және оның бірліктерінің өзара грамматикалық, семантикалық әрі логикалық байланыста болуының қажеттілігі.

Мәтіндерді өзара байланысты немесе байланыссыз деп ажыратудың негіздеріне сүйене отырып, соңғы жылдарғы зерттеулерде оның сыр-сипатын ашуға арналған талдау (анализдеу) әдісінен гөрі, жинақтау (синтездеу) әдісінің көбірек қолданылып жүргенін байқауға болады. Қиысқан мәтіндер жайлы теориялар, әсіресе сөйлемдердің өзара тіркесу заңдарының синтаксистік тұрғыдан баяндалған нақты ережелері көрсетілген қай тілге де қатысты зерттеулер, соның ішінде қазақ тіліне байланысты ізденістер әлі де болса аздық етеді.

Отандас тілші ғалымдарымыздың дәстүрлі лингвистикалық зерттеулері де жеке сөйлем шеңберінен тыс шыға қоймағандығы бәрімізге мәлім. Сондықтан мәтіннің тек қана сөйлемдер мағынасымен шектелмейтінін, одан да ірі мәтін бөлшектерінің белгілі реттілікті сақтайтын фразадан жоғары бөліктерімен тұтастық мағына түзейтінін ескеру қажет болып отыр [Гальперин, 1981]. Психологтардан кейін көп ұзамай-ақ, тілшілердің өздері де мәтін бойындағы жалпы тақырыптың **подтема, субподтема, микротема** және т.б. терминдермен аталатын шағын тақырыпшаларға ажыратылатынын қолдай бастағаны белгілі [Новиков, 1983; Маслов, 1975].

Тіл жүйесінде осы күнге дейін толық шешімі табылмай жүрген мәселелердің ұзын-ырғасы аз емес. Солардың бірі – сөйлемнің жалпы тіл жүйесіндегі алатын орны, екіншісі – сөйлем құрастырудың семантикалық және синтаксистік принциптері. Сол сияқты сөйлемді мәтін ішіндегі негізгі семантикалық компонентке жатқызу не жатқызбау мәселелері де өз шешімін таппай келе жатқан жайдың бірі [Зубов, 1987].

Мәтін бойындағы абзац мәселесі де, қазақ тілшілерін айтпағанның өзінде, шетелдік тіл мамандары арасында да әлі де болса көп талас тудырып жүргені мәлім. "Осы кезге дейін абзацты мәтіннің қай бірлігіне жатқызамыз?" деген сұраққа әртүрлі пікірлер айтылып жүр. Ғалымдардың кейбіреулері оны композициялық (құрылымдық) бірлік десе, екінші біреулері – логикалық не семантикалық топқа қосып жүр. Тағы бір зерттеушілердің айтуынша, абзац – бір орынға жай ғана топталған сөйлемдердің жиынтығы емес, ол автордың сөйлеу практикасында өз ойын оқырманның саналы түрде қабылдауына, оны меңгерудің амалдарын жеңілдетудің мақсатына лайықтап құрған мәтіндік бірлік, дейді. Мысалы, Г.В.Бондарко, Ю.А.Шрейдерлердің айтуынша, мәтінді абзацтарға бөліктеу – қабылдаушының мәтін мазмұнын дұрыс түсініп, есте қалдыру мүмкіндігін жеңілдету үшін жасалатын арнайы әрекеті екен [Бондарко, Шрейдер, 1978]. Біз де осы соңғы пікірді ұстаным ететінімізді айта кетпекпіз.

Қоршаған табиғи ортаның әртүрлі нақты нысандардан тұратындығы және оларға тән динамикалық, статикалық, құрылымдық қасиеттері болатындығы, кеңістікке тән мезгілдік қарым-қатынас қасиеттерін иеленетін жиындардан тұратыны айқын. Мұндай ортаны танып-білу дегеніміз, сол ортаны мүмкіндігінше кіші бөліктерге бөліп қарастыру, яғни ол біздің санамызда немесе мәтін ішінде кездесетін өзара мағыналық байланыстағы сөздер бірлігінің (блогінің) көрінісі екен [Зубов, 1985, 34-б.]. Сондықтан да біз, көптеген тілшілердің, психологтар мен логика мамандарының, сонымен қатар өз зерттеулеріміздің негізінде, аталған өзара мағыналық байланыста тұратын сөздер блогі – жазба мәтін, көп жағдайда, абзацпен сәйкес келеді деп ұйғаруды жөн көрдік. Жазба мәтіннің абзацтардан құралған бөліктерге бөлшектелу себебін анықтау үшін ғалымдар әртүрлі тәжірибелер жүргізіп,

олардың кездейсоқ жағдай емес екендігін, оның мәтін иесінің мазмұнды саналы түрде "тізгіндеу" қажеттігінен туындайтынын дәлелдеді.

Абзацқа енетін мәтіннен алынған сөйлемдерді өзара байланысы жоқ, бірнеше абзацтардан алынған сөйлемдер тізбегімен салыстыра қарастырғанда төмендегідей ерекшеліктері байқалған:

а) абзац құрайтын сөйлемдер тізбегі бір **микротақырыптың** не болмаса бір-бірімен тығыз байланыста тұрған **бірнеше микротақырыптың** мазмұнын анықтайды;

ә) бір абзацтың ішіндегі сөйлемдер өзара арнаулы лексикалық амал-тәсілдер арқылы байланыста болуы керек;

б) абзац ішіндегі сөйлемдердің бір-бірімен белгілі синтаксистік байланыста болуы да шарт [Агроскина, 1972].

Ғылыми-техникалық мәтіндерде бұл ережелер тұрақты қасиетке ие, яғни стереотипті сипатта болса, ал көркем әдебиет стиліндегі мәтіндерде абзацтар көрінісі біршама еркін. Соңғыларда аталған ережелер шығармашылыққа қатысты өзгеріп, әр авторға тән өзіндік айырым-белгілері болуы да мүмкін. Сол сияқты кейбір ғалымдар мәтіннің семантикалық құбылуын абзацтың құрылымдық (композициялық) ерекшелігімен де байланыстырады.

Біздің қарастыратынымыз – мәтін құрылымындағы абзацтардың атқаратын қызметіне тән белгі – стереотиптік (қалыптық) қасиет. Күнделікті өмірімізде, яғни білім алу кезеңінде не тіл үйрену жағдайында, біздің сана-сезімімізде әр стильге, яғни ғылыми, публицистикалық, көркем, т.б. әдебиеттердің негізгі белгі-қасиеттері тұрақты стереотиптік түрде жинақты орын алады. Аталған қасиеттер жиынтығын мәтіннің синтаксистік құрылымының ықтималдық моделі деуге де болады.

Сөйлеу қызметінің "астарында", яғни сөйленіс мазмұнында оның түп қазығы ретіндегі заттық мағына жатқаны айқын. Мұндай тұжырым бойынша тілдік қарым-қатынастың негізі – дискретті, яғни бөліктік тұрпатта болуы, олай болса оны сәулелендіретін мәтіннің де бөліктелген "кванты" немесе "бөлімшесі" – "логикалық қалып" немесе логикалық стереотип болады деп ұйғарған жөн [Герд, 1978, 55-58-бб.; Серкова, 1981]. Сол сияқты көптеген ғылыми ізденістерде де ми қызметі тұрпатын элементарлық ақпараттар процесінің жиынтығы деп қарастыру жиі кездеседі.

Сонымен, қазақ тіліндегі мәтіндер ішіндегі абзацтарды да "мағыналық квант" және "тұрақтанған синтаксистік үлгі" ретіндегі мәтіннің ең кіші логика-семантикалық бөлігі деп қарастыруға әбден болады.

Орыс тілі және басқа да үнді-еуропа тілдеріндегі зерттеулерді негіз етіп алып, біз қазақ тіліндегі мәтіндердің стильдік айырымдары бойынша ажыратылған абзацтарға тән функционалды-мағыналық топтарының статистикалық ерекшеліктерін анықтауға талпыныс жасадық [Жубанов, 2002]. Осы аталған зерттеуімізге қысқаша тоқталайық.

Барлық тілдердің, соның ішінде қазақ тілінің мәтіндерінің айрықша қасиеті – болмыстағы нысандар мен құбылыстардың қарым-қатынастарын сипаттайтын оқиғаны баяндайтын сөйленіс түрлерінде сөйлемдердің бір-бірімен тығыз мағыналық байланыста болуы. Ал бұл, болмыстағы кейбір заттар мен құбылыстар тізбектеріндегі уақыт не кеңістік арасында пайда болатын қатынастың мағыналық тұтастыққа ие болатындығының айғағы [Мильчин, 1980].

Ғалымдардың тұжырымдауынша, мәтін бойындағы сөйлемдерде айтылымдар мен сөйлеу типтері: суреттеу, баяндау, пайымдау түрінде көрініс табады. Ал олардың

әрбіреуі белгілі-бір функционалды-мағыналық мәтіндік және тұрақты дәрежедегі құрылымдық сипатқа ие.

Енді осы аталғандарды сөйлеу типтеріне, яғни баяндау, суреттеу және пайымдау түрлеріне ажырата білу үшін олардың анықтамаларына тоқталайық.

1-анықтама. *Егер мәтін ішіндегі фразалардың немесе одан да ірі мәтіндік бірліктердің орналасу тәртібі оқиға желісінің уақыттық деңгейіне қатысты болса, ондай сөйлеу түрі* ***баяндау типі*** *деп аталады.*

Баяндау типі іштей етістіктердің семантикасына қарай, **нақты-оқиғалық** және **жалпылама-оқиғалық** болып ажыратылады. Ал мәтіндегі хабарлар мен жалпы ақпараттық баяндаулар тізбегі үшін кейбір қимыл-әрекеттерді немесе тұрақтылық күйді баяндауда **мезгілдік нақтылықтың** қажет болмауы да мүмкін.

2-анықтама. *Егер мәтін бойындағы фразалар немесе одан да ірі мәтіндік бірліктер кеңістік, функционалдық не басқа деңгейдегі қатынаста болатын болса, ондай сөйлеу түрі* ***суреттеу типі*** *деп аталады.*

Оқиға желісіндегі **суреттеу типін** ажыратудың ең маңыздылары төменде аталған **құрылымды-мағыналық** түрлер [Мильчин, 1980, 117-б.]:

1) табиғатты суреттеудің түрлері;

2) ғимараттың ішкі құрылысын суреттеу;

3) кейбір жанды не жансыз нысанның сыртқы пішінін суреттеу. Мысалы, адамның сыртқы түрін, киімін, жан-жануардың, өсімдіктің, ғимараттың т.б. сыртқы пішінін суреттеу;

4) кейбір тірі жәндікті (адам, жануарлар) суреттеу немесе жансыз заттарға берілетін суреттеме-сипаттамалар. Мысалы, адам баласының мінез-құлқы, қызмет түрі, оның икемдік қасиеттері және т.б.

3-анықтама. *Егер фразалар немесе одан ірі мәтіндік бірліктер себеп-салдар деңгейіндегі қатынаста болса, ондай сөйлеу түрі* ***пайымдау типі*** *деп аталады.*

Функционалды-мағыналық сөйлеу типі ретіндегі *пайымдаудың* түрін логикалық тұрғыдағы ой қорытындылау (ой түйіндеу) дәрежесінде түсіну қажет, яғни логикалық ойлаудың ең жоғарғы формасы деп түсінген жөн. Басқаша айтсақ, *пайымдау типі* логикалық категорияға қатысты бола тұра, мәтіннің белгілі-бір құрылымды-мағыналық бірлігі болып та есептеледі [Мильчин, 1980, 118-б.].

Қазақ тіліндегі тәжірибе ретінде алынып отырған ғылыми-көпшілікке арналған мәтіндерді зерттеу негізінде **пайымдау типінің** алты түрі және оларға тән тілдік тұрпат анықталды.

Тәжірибелік мәтін бойынша, компьютер арқылы жүргізілген статистикалық есептеу негізінде ең жиі қолданыстағы **пайымдау түрлері** мыналар:

1) есім түріндегі бағалау, пайымдау түрі;

2) тұрақтылық күй мағынасындағы пайымдаулар;

3) нақты немесе жорамал іс-әрекетті негіздеу мағынасындағы пайымдаулар;

4) кейбір іс-әрекеттің қажеттігін, мүмкіндігін білдіретін немесе қалау не тілеу мағынасындағы пайымдаулар;

5) алдын ала келісілген іс-әрекет мағынасындағы пайымдаулар;

6) кесімді түрдегі терістеу не тұжырымдау мағынасындағы пайымдаулар.

Мұндай **баяндау, суреттеу, пайымдау типтерінің** айырымдық ерекшеліктері мәтін ішінде ең кіші семантика-синтаксистік бірлік арқылы, яғни **абзац түрінде** шектеліп барып, көрініс табады. Сондықтан біз өз зерттеуімізде мәтіндегі абзацтарды –

баяндау абзацы, суреттеу абзацы және **пайымдау абзацы** деп сараладық. Кейбір абзацтарға сөйлеу типтерінің аталған үш түрінің де қасиеттерінің қатысы болып жатса, ондайды **аралас типті** абзацтар деп атадық.

Кез келген мәтінде (ғылыми стильден басқа) сөйлеудің диалог түрі кездеседі, сондықтан сөйлеуге қатысушы екі жақтың да айтылымдарын (төлеу сөзден басқа) абзацқа жатқызуды ұйғардық. Осының салдарынан жоғарыда аталған – **баяндау, суреттеу, пайымдау абзацтары** түріне қосалқы тағы да мынадай функционалды-мағыналық түрлер ажыратылды:

а) хабарлау мәнді абзацтар;

ә) сұрақ мәнді абзацтар;

б) әсер ету (лепті мағыналы) абзацтар;

в) ниет, ойды білдіретін абзацтар;

г) пікір тудыру абзацтары;

д) аралас абзацтар [Мильчин, 1980, 118-119-бб.].

Осы аталғандардан басқа, біз өз зерттеуімізде абзацтар қызметіне, орналасу тәртібіне қарай: мәтіннің бастапқы жағында орналасқан **кіріспе** абзацтарды, ішкі бөлігінде кездесетін **медиалды** абзацтарды және мәтіндегі оқиға желісін аяқтайтын **қорытынды** абзацтарды ажыраттық.

Зерттеу нысаны ретінде үш түрлі стиль бойынша: газет мәтіні, ғылыми-көпшілікке арналған мәтіндер және көркем әдебиет мәтіндерде кездескен абзацтарды жоғарыда аталған түрлер бойынша топтап, компьютер көмегімен алынған статистикалық деректерге салыстырмалы талдау жүргізілді.

Енді осы аталған тәжірибелік сипаттағы нәтижелердің кейбіреулеріне қысқаша тоқталайық.

Көркем әдебиет бойынша *487* абзацтан тұратын *20* түрлі мәтін қарастырылған болатын. Алынған статистикалық нәтижелер арнайы кесте түрінде көрініс тапты. Сол кестелерден алынған кейбір деректерді ғана сөз етейік.

Аталған мәтіндер ішіндегі **кіріспе** абзацтар, **медиалды (ішкі)** және **қорытынды** абзацтар функционалды-мағыналық жағынан сараланып, олардың қандай типтері өнімді, қандайы өнімсіз екені сараланып, нәтижесінде көркем әдебиетке қатысты мәтіндер бойынша айқындалған *44* **кіріспе абзацтардың 30 (68%) баяндау мәнді** абзацтары екендігі анықталады. Сол сияқты, абзацтардың *20,4* пайызы **аралас типті** абзацтарға жататыны дәлелденді. Ал, **кіріспе** абзацтарда функционалды-мағыналық типтердің басқа түрлері, яғни **суреттеу** мен **пайымдау түрлері** аз қолданыс тапқаны анықталды.

Көркем әдебиет мәтіні бойынша, **баяндау типті абзацтардың ішкі (медиалды)** және **қорытынды** абзацтар ішінде де көп қолданыста болуы (*54%* және *69%*) статистикалық жолмен дәлелденді.

Сан жағынан, *404* **медиалды (ішкі)** абзацтар қатарынан жоғарыда аты аталған типтердің барлық түрлері де (біреуі аз, біреуі көп) орын тапты деуге болады. Олардың ішінде бірінші орында – **баяндау**, одан кейінгі екінші орында – **аралас типті**, сосын барып, жиілігінің кему ретімен алғанда: **әсер ету, сұрақ мәнді пайымдау** және **суреттеу мәнді** абзацтар саралау санағынан өтіп, олардың типтері анықталды. Бұларда ең жиі кездесетіні **баяндау мәнді** абзацы болғанымен (*69%*) әрі қарайғы бөлінуде басқаша, яғни екінші орында – **"әсер ету"** (*15,4%*), үшінші – **"пайымдау"**, ең соңында **"аралас мәнді"** абзацтар байқалды.

Газет мәтіні бойынша шамалас көлемдегі, тақырыптас ("Алты жастағылар мектепке барады") ***35 мақала*** зерттелді. **Бастапқы** және **қорытынды** абзацтары негізінен **"баяндау"** абзацтары ***(89%, 96%)*** болса, **ішкі (ортаңғы, медиалды) абзацтар** ішінде ең көбі – **баяндау мәнді** абзац, сосын **аралас мәнді *(17%)*, суреттеу *(9%)*, сұрақ мәнді *(2%)*** және **әсер ету мәнді *(1%)*** абзацтар типтері.

Ғылыми-көпшілікке арналған қазақ мәтіндерінің ***47-сі*** қарастырылды. Бұл аталған функционалды тип мәтінде де жоғарыда аталған жалпы заңдылық, яғни баяндау мәнді абзацтарының (кіріспе, ішкі, қорытынды) басымдылығы сақталады. Дәлірек айтсақ, бастапқы және қорытынды абзацтардың ***70*** пайызы, ішкі абзацтардың ***50*** пайызы баяндау типті болып кездескен. Бұл жерде айта кететін жайт, **бастапқы** абзацтардың ***17*** **пайызы аралас типті**, бірақ олар негізінен баяндау мен пайымдау типтеріне ұқсас келеді. Мұндай аралас типтердің басымдылығы мәтіннің ортаңғы және соңғы бөліктеріне де тән екені айқындалды. Ал, **ішкі** абзацтар тобына жататындарда **баяндау типінен** кейін екінші орында **пайымдау мәнді** абзац екені анықталды. Ол барлық **медиалды** абзацтардың ***28,4*** пайызын құрайды. Тағы да бір айта кететін жайт, ғылыми-көпшілікке арналған қазақ тіліндегі мәтіндердің "шеткілерінде", яғни "кіріспе" және "қорытындыларында" пайымдау типі үшінші орынға ауысып, ***5%*** және ***13%*** шамасында кездесті.

Бұл жүргізілген тәжірибелік зерттеуіміздің негізінде абзац түрлерінің біз көрсеткен пайыздық дәрежелері аталған стильдердегі мәтіндердің барлығына бірдей тән қасиет болмауы да мүмкін. Себебі, біздің зерттеуіміздегі қиындық – абзац түрлерін әртүрлі типтерге жатқызудың айырымдық жағын жете айыра білу, яғни олардың шекарасын ажырата білу. Бұл жерде зерттеушінің сол мәтінде сөз болатын тақырыпқа маман болуы қажет және мәтінге тән басқа да өзіндік айырым-белгілерге байланысты статистикалық нәтижелер сандық-сапалық өзгерістерге ұшырауы мүмкін.

"Қазақ тілінің мәтін туындату алгоритмін" құру мәселесінде мынандай мақсат пен міндеттер ең алдымен шешілгені жөн деп санаймыз:

1) формалды түрде мәтіннің жалпы тақырыбының берілуі;

2) мәтіннің негізгі семантика-синтаксистік бірліктерінің айқындалуы;

3) осы бірліктерді тұтастық байланыста ұстайтын ықтималды-анықталған ережелердің берілуі [Мильчин, 1980].

Бұл аталған мақсат пен міндеттер өз шешімдерін табуы үшін қазақ тіліндегі әртүрлі құрылымдағы және көлемдегі мәтін түрлері мен типтерін арнайы тәсілдермен терең зерттеуді қажет етеді.

Жоғарыда баяндалған нәтижелерді әлі де нақтылай түсу тұрғысынан абзац түрлерінің мазмұндық жағынан заттық-логикалық топталуы мәтіннің "туындау" мәселесінде аса зор орын алады.

7.2 Мәтін абзацтарының заттық-логикалық мазмұны

Ана тілімізді зерттеу әдістерін заманға сай дамыту үшін, оны бейнелейтін мәтін бірліктерінің сыр-сипатын, олардың өзара байланыстық мәнін және белгілі бір тақырыпқа қатынасын ашу мәтін лингвистикасының өзекті мәселелерінің бірі екендігі жоғарыда айтылған болатын. Осымен қатар автордың мәтінді абзацтарға бөлшектеудің заңды құбылыс екендігі және мәтін абзацының синтаксистік және

мазмұндық құрылымы моделінің элементі болатындығы да сөз болды. Абзацқа тән қасиеттерді саралаумен бірге, олардың мәтін құрылымына қатысты – кіріспе, медиалды (ішкі), қорытынды деп айырым табуымен қатар, мазмұндық жағынан – **баяндау, суреттеу, пайымдау** деп аталатын функционалды-мағыналық топтарға ажыратылатыны дәлелденді [Жубанов, 2002, 118-124-бб.; Зубов, 1989, 7-б.].

Сонымен, абзац мазмұны арқылы ақиқат болмыстағы шекті нысандардың қасиеттері мен олардың өзара қатынастық мүмкіндігі баяндалатын микрожағдаят берілетіні белгілі болды. Мәтін абзацтарын функционалды-мағыналық топтарға ажырата зерттеу арқылы біз тек мазмұнның жалпы типтік жақтарын ғана анықтай аламыз да, ал оны нақтылай түсудің жолы абзац мазмұнының заттық-логикалық түсінігіне тікелей қатысты [Зубов, 1989; Гришина, 1982].

Мәтін құрылымының статикалық және динамикалық құрастырушыларының болуын қолдауымызға қатысты ақиқат болмыстағы қандай нысандар функционалды-мағыналық топтарының қайсысына қатысты екендігі және олар қандай тірек сөздер арқылы берілетіндігін зерттеу – аса маңызды мәселе.

Осы айтылғандарды ескере отырып, әрбір функционалды-мағыналық типке қатысты "абзацтың" заттық-логикалық мазмұнының анықтамалары төмендегідей [Жубанов, 2002, 125-126-бб.; Зубов, 1989, 7-12-бб.] :

1-анықтама. *Мәтін абзацтарында дәйекті түрде баяндалатын ақиқат болмыстағы іс-әрекет пен қалыпты жағдайлардың егжей-тегжейіне дейін тәптіштелген жанды не жансыз нысандарды орын мен уақытты айқындауына қарай жіктелген жалпы мәтін мазмұнына қатысты негізгі тірек сөздер (НТС) мен екінші дәрежелі қосалқы тірек сөздер (ҚТС) жиынтығын баяндау абзацтарының заттық-логикалық мазмұны деп атаймыз.*

2-анықтама. *Мәтін абзацтарында сөз болатын кішігірім оқиғадағы негізгі және екінші дәрежелі жанды мен жансыз нысандарды орын мен уақытты суреттеуіне қарай сараланып, жалпы мәтін мазмұнына қатысты ақиқат болмыстық деректері (белгілер, құбылыстар, оқиғалар, көзқарастар, теориялар* және т.б.*) түгелдей саналып аталынатын негізгі тірек сөздер (НТС) мен екінші дәрежелі қосалқы тірек сөздер (ҚТС) жиынтығын суреттеу абзацтарының заттық-логикалық мазмұны деп атаймыз.*

3-анықтама. *Мәтін абзацтарында баяндалатын кішігірім оқиғадағы негізгі және екінші дәрежелі жанды не жансыз нысандарға орын мен уақытқа байланысты ой тізбегінің жалпы мәтін мазмұнына қатысты қатынастарын анықтайтын негізгі тірек сөздер (НТС) және екінші дәрежелі қосалқы тірек сөздер (ҚТС) жиынтығын пайымдау абзацының заттық-логикалық мазмұны деп атаймыз.*

Осы анықтамалардың негізінде қазақ тіліндегі әртүрлі стильдерді анықтайтын мәтін абзацтарын, олардың функционалдық-мағыналық және құрылымдық (кіріспе, медиалды, қорытынды) типтері бойынша ажырата отырып, заттық-логикалық мазмұнына қарай топтап, олардың ішкі типтерін ажырату арқылы статистикалық зерттеу нәтижелерін қысқаша түрде атап өтейік.

Мұндай топтау жұмыстары, яғни абзацтарды функционалды-мағыналық жағынан ажырата білу зерттеушіге аса ұқыптылықты қажет ететіні анық. Қазақ тілінің үш түрлі стилі (ғылыми-көпшілікке арналған стиль, газет стилі, көркем әдебиет стилі) бойынша аталған зерттеу жұмыстары алғашында абзацтардың мәтін бойында орналасу тәртібіне қарай, яғни құрылымдық тұрғыда қарастырылды.

Ғылыми-көпшілікке арналған ***47*** мәтін (Көшербаев К. Минералдар әлемінде. Алматы, 1986; Қалыбеков Қ. Экология және ашық кен. Алматы, 1988) бойынша жүргізілген зерттеу жұмысының нәтижелерін автордың мына ғылыми еңбегі бойынша [Жубанов, 2002, 126-133-бб.] толығырақ танысуға болады. Аталған ***47*** мәтіндерде бастапқы мен кіріспе абзацтар саны бір-біріне тең (және ол – ***93)***. Жоғарыда келтірілген анықтамаларға сай, мәтіндерді заттық-логикалық мазмұнға талдау нәтижесінде ***20*** типтік тұжырымдаулар топталып алынып, топтамалардың статистикасы анықталды және оларды шартты белгілермен кодтау принципі шешілді. Аталған зерттеу нәтижелері арнайы кесте арқылы көрініс тапты. Аталған кестелер *5* бағанадан тұрады:

1 – типтердің қатар саны;
2 – бастапқы абзацтар типінің тұжырымдалуы;
3 – типтің коды (шартты белгісі);
4 – типтің абсолютті жиілігі;
5 – типтің қатынастық жиілігі.

Бастапқы абзацтардың заттық-логикалық мазмұны тұрғысынан топталған типтері қолдану жиілігі жағынан біркелкі емес болуынан, олардың жоғары жиіліктегі, яғни ***"9"*** және одан жоғары дәрежелері ғана алынып, *1-кесте* үзіндісі түрінде беруді ұйғардық. Кестеде көрсетілген тұжырымдаулар, біз қарастырған ғылыми-көпшілікке арналған мәтіндерге тән мазмұндық типтер деуге болады. Мазмұн – тұжырымдауларда аталған бас нысан төңірегінде шоғырланып, сол нысанның мәнін ашуға қатысты автордың (зерттеушінің) өткен және осы шақтарда істелген әрекеттері мен болар шақта іске асатын әрекеттерінің нәтижелері ретінде сөз болады.

Мысал ретінде, осындай ғылыми-көпшілікке арналған – ***"Кенді ауыр суда байыту"*** тақырыбы бойынша жазылған мәтіннің кіріспе абзацтарының біреуін ғана қарастыралық: *"Тығыздығы 1кг/м³-ден асатын кез келген зат суда батады. Мұндай заттардың өте ұсақ түйіршіктерінің өзі суда қалқуы мүмкін емес. Ал мына көріністің мәні ерекше: жұдырықтай кесектер суда қалытқыдай қалқуда... Мұның мәні – жай судың орнына салмағы ауыр су алынуында...".*

Бұл абзацты *1-кесте* үзіндісіндегі тұжырымдаудың төртінші санына жатқызуға болады деп санаймыз, яғни бас нысан – тығыздығы ***1кг/м³***-ден асатын ***заттың*** суға бататындығын айрықша атап көрсету. Бұдан соң, бас субъектінің – тәжірибе иесінің (ғалымның) бас нысанды суға батқызбау, яғни оның су бетінде қалқып жүруі үшін істелетін әрекеттері – судың тығыздығын ***1кг/м³***-ден де жоғары асыру. Сонымен бірге, болатын (болған) жайтты қайталап атау (констатация фактов) ретінде қысқаша түрде баяндалуы.

1-кестеде көрсетілгендей, осы мазмұнға формалды түрде сай келетін абзацтарды ***"M004"*** деген шартты белгімен кодтау ұйғарылды. Кестедегі осы тәріздес белгілер мәтін абзацтарының заттық-логикалық мазмұнын формалдау ісіне және солар арқылы мәтіннің қысқаша түрде мазмұнын ашу кезінде аса маңызды рөл атқарады деп санаймыз. Осылайша, абзац мазмұнын заттық-логикалық тұрғыдан талдаулар ғылыми-көпшілікке арналған мәтіннің ***353*** **медиалды (ішкі, ортаңғы)** абзацтары бойынша да жүргізіліп, соның негізінде саны ***23-ке*** тең типтік тұжырымдаулар ажыратылды. Біз олардың жиі қолданыс тапқан ***5 типін*** ғана *2-кесте* үзіндісі арқылы көрсетуді жөн көрдік.

Мәтіннің ортаңғы абзацтарындағы **бас субъекті** мен **автордың** (тәжірибешінің) кейбір іс-әрекеттерінің (жүріс-тұрыс, бұлшық ет жұмысы мен қызметі) талдау арқылы бас нысанды танып-білу процесі іске асты.

1-кесте (үзінді). Ғылыми-көпшілікке арналған мәтіндердің кіріспе абзацтарын заттық-логикалық мазмұнға талдау нәтижесі

№ к/с	Кіріспе абзацтар типтерінің тұжырымдалуы	Тип коды	Абсолютті жиілік (Fi)	Қатыстық жиілік (fi)
1	2	3	4	5
3	Бас нысанның басқа да бірінші және екінші дәрежелі нысандарға қарағанда өзіне ғана тән ерекшелігін қайталап атау (констатация фактов).	***МОО3***	**11**	**0,1183**
4	Бас нысанға ерекше көңіл аударуды айырықша атап көрсету және оған байланысты автордың (тәжірибешінің) іс-әрекетін қысқаша қайталап атау.	***МОО4***	**13**	**0,1398**
6	Кейбір нысандардың мәнін пайымдау (мақұлдау).	***МОО6***	**10**	**0,1075**
7	Автордың болашақ іс-әрекетіне байланысты шарттарды сөз ету.	***МОО7***	**10**	**0,1075**
11	Автордың кейбір бас нысанға қатысты жай ғана пайымдауларын қайталап атау.	***МО11***	**9**	**0,0968**

Ғылыми көпшілікке арналған мәтін бойынша талданған медиалды абзац ретінде ***(Н10-08)*** мәтінінен мысал: *"Магнитті байытудың өнімділігін арттырудың бір жолы – магниттілігі аз минералдардың магниттік қасиетін алдын ала күшейту. Мысалы, гематит, сидерит және лимонит секілді минералдардың магниттілігі әлсіз. Егер осы минерал рудаларын 550-600 градуста көміртегін қосып күйдірсе күшті магниттеледі".*

2-кесте (үзінді). Ғылыми-көпшілікке арналған мәтіндердің медиалды абзацтарын заттық-логикалық мазмұнға талдау нәтижесі

№ к/с	Кіріспе абзацтар типтерінің тұжырымдалуы	Тип коды	Абсолютті жиілік (Fi)	Қатыстық жиілік (fi)
1	2	3	4	5
3	Бас нысанның басқа да бірінші және екінші дәрежелі нысандарға қарағанда өзіне ғана тән ерекшелігін қайталап атау (констатация фактов).	***МОО3***	**11**	**0,1183**
4	Бас нысанға ерекше көңіл аударуды айырықша атап көрсету және оған байланысты автордың (тәжірибешінің) іс-әрекетін қысқаша қайталап атау.	***МОО4***	**13**	**0,1398**
6	Кейбір нысандардың мәнін пайымдау (мақұлдау).	***МОО6***	**10**	**0,1075**
7	Автордың болашақ іс-әрекетіне байланысты шарттарды сөз ету.	***МОО7***	**10**	**0,1075**
11	Автордың кейбір бас нысанға қатысты жай ғана пайымдауларын қайталап атау.	***МО11***	**9**	**0,0968**

Заттық-логикалық мазмұн жағынан бұл медиалды (ішкі) абзац *2-кестедегі* қатар саны *10*-ға тең тұжырымдауға сай келеді деп есептеуге болады, яғни "Басқа бір екінші дәрежелі нысандарға қатысты кейбір шарттарды қайталап атау". Ал мәтіндегі мұндай тұжырымдауға бағынатын абзацтардың қолданылуының абсолютті жиілігі ***F=37***, ал қатынастық жиілігі ***f=0,1048*** немесе олар орта есеппен мәтіндегі барлық абзацтың ***10,5*** пайызын құрайды.

Келесі қарастыратынымыз – ғылыми-көпшілікке арналған ***47*** мәтіндер ішінен, біздің ұйғаруымызша алынған, ***57*** қорытындылау абзацтары. Зерттеу нәтижесінде мұндай абзацтардың ***10*** түрлі типтік тұжырымдау топтары ажыратылды және олардың статистикасы мен кодтау принциптері анықталады *(3-кестенің* үзіндісі).

Ғылыми-көпшілікке арналған мәтіндер бойынша қорытынды абзацтарға қатысты тұжырымдаулардың кестеде көрсетілгендей мағынада берілуі өзінен-өзі түсінікті сияқты, себебі көп жағдайларда мәтін соңындағы абзацтарда болған іске қорытынды көзқарас айтылып, автордың болашақтағы іс-әрекетінің бағыт-бағдары белгіленеді.

Енді осы айтылғандарға мысал ретінде функционалды-мағыналық сипаты бойынша баяндау типіне жататын, ал заттық-логикалық мазмұн жағынан (біздің ұйғаруымызша) *3-кесте* үзіндісіндегі рет саны ***8-ге*** тең типтік тұжырымдауға жататын абзац мәтінін келтірейік: *"Атмосфера ауасының тазалығы жолында күрес жүргізу – ең алдымен адамдардың денсаулығын сақтау мәселесі. Сондықтан адамдардың денсаулығына бірінші дәрежелі маңыз аударылатын елде – атмосфера ауасының гигиенасы гигиеналық ғылымның арнаулы саласы ретінде құрылып, ойдағыдай дамып келеді".*

Бұл абзацтағы заттық-логикалық мазмұн бойынша бас нысан – ***атмосфера ауасы***. Бірінші сөйлемде автор *таза атмосфералық ауа* мен *адам баласының денсаулығы* арасындағы тығыз байланыстың барына аса көңіл аудара келе, бас нысанға, яғни ***атмосфералық ауаға*** сипаттама беруге талаптанады. Абзацтың екінші сөйлемінде бұл мәселені ғылыми тұрғыда қарастыру қажеттігін және оның біздің елде ойдағыдай дамып келе жатқанын сөз етіп қорытындылайды.

Сонымен, осы зерттеу нәтижелері негізінде айтарымыз – қазақ тіліндегі ғылыми-көпшілікке арналған мәтін мазмұнының дұрыс ашылуы оның құрылымдық жағына байланысы бола бермей-ақ, заттық-логикалық мазмұн тұрғысынан қарастырғанда, осы стильге тән жалпы заңдылық сақталады және ұқсас типтік тұжырымдаулар мәтіннің әр бөліктерінде де қолданыс таба алады.

Жоғарыда сипатталған әдістер негізінде мазмұндық типтерді ажырату жұмысы газет мәтіндері бойынша да жүргізілді. "Қазақстан мұғалімі" газетіндегі *"Алты жасарлар мектепке барады"* атты мәліметке қатысты (Алматы, 1986 ж.) жарияланған жалпы көлемі ***502 абзац*** құрайтын тақырыптас ***35 мақала*** компьютер арқылы заттық-логикалық мазмұн тұрғысынан статистикалық тәсілмен зерттелді [Жубанов, 2002, 133-140-бб.].

Абзацтардың мәтін бойындағы орналасу тәртібі (кіріспе, медиалды, қорытынды) ескеріле отырып, олардың типтік тұжырымдаулары *3* түрлі ***кесте*** түрінде көрініс тапты. Осы аталған зерттеу нәтижелеріне қысқаша тоқталайық.

Газет стилі бойынша *35* мәтін ішінен ***25 бастапқы*** (кіріспе) абзацтар талданып, ***11 мазмұндық типтің*** мәтін ішіндегі қолдану жиіліктері есептелді және әрбір типке қатысты **шартты белгі-код** қою принципі шешілді.

3-кесте (үзінді). Ғылыми-көпшілікке арналған мәтіндердің қорытынды абзацтарын заттық-логикалық мазмұнға талдау нәтижесі

№ к/с	Қорытынды абзацтар типінің тұжырымдалуы	Тип коды	Абсолютті жиілік (Fi)	Қатыстық жиілік (fi)
1	2	3	4	5
1	Бас нысанның мәні жайлы қорытынды нәтижелерді пайымдау	**М201**	**19**	**0,3333**
2	Бас нысанның ерекшеліктері жайлы автордың (зерттеушінің) пайымдауы	**М202**	**10**	**0,1754**
3	Бас нысанға қатысты автордың өткен шақтағы іс-әрекетін қайталап атау	**М203**	**3**	**0,0526**
4	Кейбір мәселелердің қойылуын қайта пайымдау және автордың болашақ іс-әрекетін сипаттау	**М204**	**2**	**0,0351**
5	Бас нысанға қатысты автордың болашақ іс-әрекетін қайталап атау	**М205**	**1**	**0,0175**
6	Бас субъектіге қатысты автордың қорытынды нәтижелерінің суретке (графикке, кестелерге және т.б.) байланысты пайымдалуы	**М206**	**1**	**0,0175**
7	Автордың алда тұрған іс-әрекеттерін қайталап атау	**М207**	**1**	**0,0175**
8	Мәселенің қойылуы мен оның бағалануы тұрғысынан бас нысанға қайсыбір сипаттамалардың берілуі	**М208**	**13**	**0,2283**
9	Бас нысанға қатысты екі жақты сипаттаманың қысқаша түрде берілуі	**М209**	**4**	**0,0702**
10	Бас нысан мен екінші дәрежелі нысандардың болашақ жағдаяты жайлы болжау айту	**М210**	**3**	**0,0526**
		Барлығы:	**57**	**1,0000**

Газеттік стиль бойынша ***11* заттық-логикалық мазмұн** типтерінің тұжырымдаулары ішінен ең жиі қолданыс тапқандары мыналар:

- "басқа бір жанды не жансыз нысандарға байланысты табиғатты немесе сырт жағдайды суреттеу";
- "басқа бір нысанға сипаттама беру және басқа бір субъектінің болашақ іс-әрекетіне болжау айту";
- "басқа бір нысанның мәні жайлы автордың пайымдауы және оның бас субъектіге қатысты қалпын қайталап атау";
- "бас нысанға байланысты қайсыбір сырт жағдайды қайталап атау".

Газет стилі бойынша ***175 медиалды абзац*** ішінен әртүрлі **заттық-логикалық мазмұндағы *41* типтік** тұжырымдаулар бөлініп алынды және олардың әрбіреуіне **М401-М441** аралығын қамтитын шартты белгі-кодтар сәйкестікке қойылды. Бұл кодтар газет стилін формалды пішінде зерттеу кезінде қажеттік етеді. Мысал ретінде, типтік тұжырымдаулардың кейбір жиі қолданыс тапқандарын келтірейік:

- "басқа бір жанды не жансыз нысандарға байланысты автордың кейбір деректерді қайталап атауы";
- "бас субъектіге қатысты қайсыбір іс-әрекет жүргізілгенін қайталап атау";
- "бас нысанға қатысты табиғатты немесе сырт жағдайды суреттеу";

- "бас субъектінің іс-әрекетін қайталап атау және бас нысанға сипаттама беру";

Абзацтарға мазмұн тұрғысынан заттық-логикалық талдау газет мәтінінің ***45* қорытынды абзацтарына** да жүргізілді.

Зерттеу нәтижесінде типтік тұжырымдаудың ***11*** түрі топтарға бөлініп, қажетті деген санақ жұмыстары іске асты. Мысалы, мына тұжырымдауларды жеке бөліп атауға болар еді:

- "бас субъектіге қатысты автордың болжам айтуы";
- "бас және екінші дәрежелі субъектілер (жанды, жансыз нысандар) жайлы автордың кең түрдегі пайымдаулары";
- "мәтінде сөз болған оқиға жайлы автордың қорытынды пайымдау деректерін қайталап атау".

Мысалы, ***"Балабақшада оқиды"*** атты мақала мынадай қорытынды абзацпен аяқталған: *"Алты жастағыларды оқытуда тек мектепке дейінгі мекемелерге арқа сүйеп жайбарақаттық танытпай, мектептердің де мүмкіндіктерін молайту қажет. Бұл мәселе облыстық және қалалық халық ағарту бөлімдері басшыларынан барынша жауапкершілікті талап етеді".*

Бұл абзацтағы заттық-логикалық мазмұнға көңіл бөлсек, бас объектіден ***(мектепке дейінгі мекеме)*** басқа, мектептің өз мүмкіндігі, облыстық, қалалық халық ағарту бөлімдері мен олардың басшыларының да (жанды, жансыз нысандар) мүмкіншіліктері де сөз болады. Сондықтан да, ең ықтималды типтік тұжырымдалу – **"бас және екінші дәрежелі жанды және жансыз нысандар жайлы автордың кең түрде пайымдауы"** типі сай келеді деп ұйғарған жөн сияқты.

Газет мәтіннің құрылымындағы барлық абзацтарды ғылыми-көпшілікке арналған стиль абзацтарымен салыстыра қарастырғанда кіріспе, медиалды, қорытынды абзацтардың заттық-логикалық мазмұн тұрғысынан және құрылымдық жағынан қарастырғанда өздеріне тән ерекшеліктермен бірге, ортақ тұстары да кездеседі. Сол сияқты ғылыми-көпшілікке арналған мәтін өзінің стереотиптік қасиетімен де ерекшеленетінін байқауға болады.

Жоғарыда сипатталған әдістер негізінде ***20*** түрлі көркем әдебиет мәтіні заттық-логикалық мазмұн тұрғысынан типтік топтау процесінен өтті. Мұндай топтау жұмысы абзацтардың құрылымдық орналасу тәртібіне қатыссыз жалпы саны ***487-ге*** тең абзацтар мазмұны кішігірім оқиғаларды бас және екінші, үшінші, т.б. дәрежелі жанды немесе жансыз нысандардың рөлдерін ескеру арқылы іске асты [Жубанов, 2002, 140-143-бб.].

Осындай заттық-логикалық талдау мен статистикалық зерттеу жұмысының негізінде ***487* абзац** ішінен ***32*** типтік тұжырымдауды бөліп алуға мүмкіндік туды.

Төменде осы типтік тұжырымдаулардың тек жоғары жиіліктегілерін ғана келтірдік:

- "табиғатты немесе сырт жағдайды суреттеу және бас субъектінің басынан өткен іс-әрекетін қайталап атау" ***(M601, F=54; f=0,1109);***
- "табиғатты суреттеу және автордың іс-әрекетін немесе қалпын қайталап атау" ***(M604, F=31; f=0,0636);***
- "бас нысан мен басқа да бас субъектілерге қатысты негізгі бас субъектінің іс-әрекетін суреттеу" ***(M605, F=40; f=0,08);***
- "автордың болашақ іс-әрекетін қайталап атау" ***(M609, F=40; f=0,0821);***
- "авторға мінездеме беру және оның іс-әрекетін қайталап атау" ***(M624, F=23; f=0,0472);***
- "екінші дәрежелі субъектінің іс-әрекетін қайталап атау және авторға мінездеме беру" ***(M625, F=20; f=0,0411);***

- "бас субъектіні іс-әрекетке талаптандыру және автордың кейбір іс-қимылы мен қалпын қайталап атау" ***(M626, F=31; f=0,0636);***

- "бас субъектіні іс-әрекетке талаптандыру және басқа да бас субъектінің іс-әрекетін қайталап атау" ***(M627, F=20; f=0,0411).***

Мысал ретінде, С.Сейфуллиннің ***"Бандыны қуған Хамит"*** атты әңгімесінен бір абзацты қарастырайық: *"Күн шайдай ашық, жылы еді. Аспанда бұлт жоқ, жел де жоқ. Айнала қалың иірім-иірім орман. Бойлап, таласа, жарыса шыққан қайың мен қарағай. Және әредік қалың қайың, қарағаймен жарыса шыққан теректер де көрінеді. Қайың мен терек ағаштардың жапырақтары салғырт тартып, сарғая бастаған. Осы орман ағаштардың ішімен, ескілеу жолмен Хамит жүріп келеді".*

Бұл абзац функционалды-мағыналық сипаты бойынша суреттеу абзацына жатады, құрылымдық жағынан – кіріспе абзацтардың бірі, ал заттық-логикалық мазмұн тұрғысынан алғанда тұжырымдау типінің ***M601*** кодына жататын, қолданылу жиілігі бойынша барлық абзацтардың ***11*** пайызын құрайтын: "Табиғатты немесе сырт жағдайды суреттеу *(көркем орман іші)* және бас субъектінің басынан өткен іс-әрекетін *(Хамиттің орман ішімен жүруі)* қайталап атау".

Көркем әдебиет стилі бойынша топталған ***32*** типтегі тұжырымдауларды талдай келе байқағанымыз, олар көбіне табиғатты, қоршаған ортаны, сырт жағдайды суреттеу, бас кейіпкердің қалыпты жағдайды немесе іс-қимылын баяндау және бас субъекті мен объектіге мінездеме, сипаттама беру, автор мен басқа да субъектілердің өткен, осы және болар шақтарындағы іс-әрекеттерін қайталап, болжап атау және т.б. Көркем әдебиет мәтіндері заттық-мағыналық тұрғысынан публицистика стилінен біршама, ал ғылыми мәтін стилінен айтарлықтай айырмашылығы болатыны байқалады.

Бұл зерттеу нәтижелері, яғни әр стиль бойынша ұсынылған типтік тұжырымдау топтары мәтін сипатына қарай, сандық не сапалық жағынан өзгеріске ұшырауы мүмкін, біз тек бұл мәселенің принциптік жағын ғана қарастырдық деуге болады.

Сонымен, мәтін абзацтарын функционалды-мағыналық мазмұнға топтау болсын не заттық-логикалық мазмұн тұрғысынан типтік тұжырымдау топтарына ажыратуға арналған зерттеулер болсын, бәрінің мақсаты қазақ тіліндегі мәтін мазмұнын формалды әдіс-тәсілдермен ашу қажеттілігінен туындайды.

7.3 Мәтіннің негізгі және факультативті мазмұнын ажырату мәселесі

Мәтін сырын ашуға арналған әртүрлі зерттеулер негізінде оның жасалуы мен мағынасы тұрғысынан қарастырғанда көп қырлы құрылым екендігі анықталды. Бұлайша тұжырымдауға себеп – кез келген мәтіннен әлеуметтік, жекелік, детерминдік және ықтималдық тереңдік және үстірт, жазба тілі мен сөйлеу тілінің (речевое) сипатын байқауға болады [Жубанов, 2002; Проблемы ..., 1983, 43-б.].

Осындай қарама-қарсы құбылыстарға мәтінге тән статикалық және динамикалық аспектілерді де жатқызуға болады. Бұл жердегі "статика" мен "динамика" терминдерін зат не құбылыстың "тыныштық" және "қозғалыс" қалыптары мағынасында түсіну қажет [Гальперин, 1981, 19-б.]. Мысалы, мәтін өзінің туындау, қабылдану мен түсіну процестері тұрғысынан алғанда, ұдайы қозғалыста немесе динамикалық қалыпта, ал одан басқа кезеңдерде лексикалық бірліктердің тізбегі ретінде және ойлау мен сөйлеу қызметі жемісінің статикалық

немесе "тыныштық" қалыптағы жағдайларда болады деп саналады [Гальперин, 1981, 19-б.; Нелюбин, 1983, 31-б.].

Толық пішіндегі сөйлемді түсінуде қабылдаушы ең алдымен, оның семантикалық мазмұны жайлы ақпаратқа, ал сосын барып, оның синтаксистік құрылымына көңіл аударады. Яғни семантикалық аспект синтаксистікке қарағанда бірінші болып қабылданады деп те тұжырымдауға болады [Леонтьев, 1969, 89- б.].

Осыған байланысты тілшілер әр мәтіннің өзіне ғана тән сөздігі мен синтаксистік құрылымы болулары қажеттігінің дәлелі ретінде мәтінге қатысты **"мәтін сөздігі", "мәтін синтаксисі"** деген терминдерді жиі түрде қолдана бастады [Проблемы ..., 1983, 56-б.]. **"Сөздік"** пен **"синтаксис"** терминдеріне зерттеушілердің әртүрлі мағынада түсінуін ескере келе, біз **"сөздік"** ұғымына байланысты – **"мазмұнды сөздер"** немесе **тірек** және **дерек сөздер** тобын есептеу қажет деген көзқарасты ұстаным етеміз [Проблемы ..., 1983, 44-б.].

Сонымен, **мәтін статикасы** дегеніміз шындық болмыста кездесетін заттар мен құбылыстар және басқа да деректерді мәтін ішінде түгелімен қамтып, бейнелеуші **"мазмұндық" сөздер жиыны**, ал **мәтін динамикасы** – осы жиын элементтері арасындағы қатынастарды бейнелеуші **синтаксис**. Аталмыш қатынастар мәтін иесінің мақсат-бағдарына, мәтін типіне, сөз шеберлігіне және басқа да факторларға қатысты іске асады. Мәтіннің статикалық және динамикалық аспектілерін растауға және оларды айыра білудің мазмұнды дұрыс түсінуге қажет екендігін дәлелдеуге арналған көптеген зерттеу тәжірибелерінің нәтижелері де бар [Лингвистические ..., 1983, 120-б.].

Мәтін мазмұнының құрылымына тән мұндай екі құрастырушының мәтін туындауы тұрғысынан алғанда **"тұтастық сипатта"** деп қарастыру көзқарасына қысқаша тоқталайық.

Бізді қоршаған ортадағы ақиқат болмыста орын алатын шағын оқиғалар мәтін ішінде шағылыс табады, яғни олар сөздер арқылы бейнеленеді. Ал адам, осы шағын оқиға элементтерін қайсыбір ықтималдық жағдайда қабылдайды екен. Әрине мұндай процесс жеке қабылдаушының өмір тәжірибесі мен тілдік деңгейінің көптеген факторларына тығыз байланысты болып келеді [Гальперин, 1981, 40-б.; Борель, 1964, 81-б.]. Осының нәтижесінде мәтінді түзуге аса құнды деп саналатын және мазмұнды ашуға қажеттіктен туындайтын референттер жүйесінің басты элементтерін таңдап алу процесі іске асады. Мұндай басты элементтер әрі қарайғы мәтінді құруға қажетті басқа да референттерден таңдалып алынатын қосалқы элементтердің ретін тағайындауға ыңғайлы жағдай туғызады [Нелюбин, 1983, 70-71-бб.].

Егер референттердің орнына мәтін ішіндегі әртүрлі атаулардың алынатынын ескерсек, онда мәтін мазмұнының формалды құрылымы – ол, материалдық белгілердің сызықтық тізбегінің формалды құрылымы немесе сөздер тізбегі екеніне көзіміз жететіні анық [Нелюбин, 1983, 28-б.].

Мәтіннің өне бойынан кездестіруге болатын және оның мазмұнын мағыналық бірлікте ұстайтын сөздерді мәтін лингвистикасына арналған орыс тіліндегі әдебиеттерде **"опорные слова"** (тірек сөздер) деп және бұл терминнің көптеген басқаша синонимдік атаулары да кездеседі: ***"ключевые слова", "смысловые вехи", "топики", "дескрипторы", "ядра", "фокальные точки"*** және т.б. [Новиков, 1983]. Әдебиеттерде мәтіннің мағыналық бірлігін сақтайтын **"тірек"** сөздер мен сөз тіркестері арқылы формалды бейнелеуді неше түрлі **"графтар құру"** жолымен сипаттау да кездеседі. Бірақ мұндай әдіс бойынша тағайындалатын формалды қарым-қатынас (байланыс) мәтінді толық қамтитын графтың денотатты құрылымын

бейнелеу мәселесіне ойдағыдай шешім болмай отыр. Бұл жағдайды мәтіннің көрінісі ретіндегі графтың барлық ұштары арасындағы ақырғы байланыс мәтінге қатыссыз, яғни мазмұндық деңгейдегі басқа ақпарат арқылы іске асатындығымен түсіндіруге болар еді [Нелюбин, 1983, 149-150-бб.].

Мәтін ішінен **"тірек сөздерді"** бөліп алып, талдау жасау негізінде олардың негізгі, қосалқы және басқа да дәрежелі сөздері арқылы хабарламаның **"жинақталған"** негізгі мән-мағынасын ашуға болатындығы анықталды [Нелюбин, 1983, 147-б.; Лингвистические ..., 1983, 123-б.]. Мәтін ішінде мағынасы тек контекстік қоршалуы арқылы ғана ажыратылатын синонимдік сөздердің кездесуі, яғни бір ғана референт бірнеше синонимдік сөздер арқылы берілуі де мүмкін [Лингвистические ..., 1983, 138-б.].

Мәтін мазмұнына қатысты мағыналық бірлікті сақтайтын әр дәрежелі контекстік тірек сөздердің сызықтық тізбегі бойынша орналасуы – олардың әр абзацтарда кездесуі мүмкіндігінің айғағы. Абзацтардың мағыналық құрылым мәселесі, шындығында, семантика-синтаксистік проблемаға тікелей қатыстығына қарамай, бұл мәселенің мағыналық жағының синтаксистік ерекшелігі. Оның лексикалық құрылымының егжей-тегжейін тәптіштеп түсінбейінше болмайтындығын баса айтқан жөн.

Абзацтарда мәтін бойындағы бірінші және екінші дәрежелі референттердің оқиғадағы мәні жағынан олардан төмен сатыда тұратын элементтермен байланыстары және қатынастары көрініс табады. Мұндай төмен дәрежелі элементтер деп отырғанымыз – әртүрлі іс-әрекет пен қалыпты жағдайлардың типтері, іс-әрекеттің болған орны мен уақыты, мақсаты және олардың себебін бейнелейтін тірек сөздер [Агроскина, 1972, 53-54-бб.].

Мәтіннің екінші құрылымдық типі – мәтін динамикасы. Ол шындық болмыстың мәтінде суреттелген белгілі-бір уақиғасына байланысты нысандар (объектілер) арасындағы қарым-қатынасты сипаттайды. Бұл жерде, тек қана мәтін ішінде атаулары тірек сөздер арқылы аталған референттерге ғана қатысты уақиғалар сөз болып, солар ғана тіркеледі [Маслов, 1975, 324-б.].

Ақиқат болмысты фрагменттерге (үзінділерге) бөліктеу кездейсоқтық жағдай емес, ол адамзаттың ұзақ уақыт бойы әлеуметтік іс-әрекет нәтижесінің үлгісіне сай қалыптасады. Сондықтан да танымдық процесі кезінде адам баласының сана-сезімі бір затқа тән қасиеттер тізбегі ішінен өз санасынан орын алған бұрыннан таныс категорияға ортақтығына қарай, болған оқиғаның кейбір үзіндісін бөліп алады [Рылова, 1973, 41-42-бб.].

Әртүрлі ғылыми бағыттағы ғалымдардың байқауынша, адам баласының танымдық процесі кезінде ақиқат шындықтың аса маңызды деген және жиі қайталанатын қатынастары адам есінде синтаксистік құрылым үлгісінде сақталады. Мұндай синтаксистік құрылымдар дегеніміз ақиқат болмыстың типтік байланыстары мен қатынастарының адам ойындағы бейнесінің қайсыбір тұрақталған және жадында жинақталған логикалық жасалымдары [Лингвистические ..., 1983, 103-104-бб.]. Басқа сөзбен айтқанда, бұл – болмыста кездесетін айтылым үлгісі немесе тілдегі **"логикалық көшірме бейне"**. Мұндай тілдегі **көшірме үлгілер жиынтығы** біршама шектелуі және әрбір қалам иесінің не шағын түрдегі салалық аяның мәтіндеріне тән қасиетте кездесіп отырады.

Соңғы жылдарғы зерттеулер нәтижесі бойынша, жазба мәтіндердің абзацтарын да, болмыстың осындай тілдік **"логикалық көшірме бейнесі"** деп қарастыруға болатындығы жиі айтылып жүр. Бұлайша ұйғарым жасау тек қана лингвистикалық

зерттеулердің нәтижелерімен шектелмей, сонымен бірге психолингвистер тәжірибелерінен де туындайды.

Мазмұнға қатысты **"тірек сөздерді"** мәтін бойынан ажыратып алудың неше түрлі әдістері бар. Бұл – мәтіннің негізгі және факультативті мазмұнын ажыратуға қатысты инженерлік лингвистиканың ең маңызды мәселелерінің бірі болып саналады. **Негізгі тірек сөздерді** (НТС) мәтін ішінен бөліп алу тәсілдерінің ішінен, қазақ мәтіндеріне ыңғайлы келетіні А.В.Зубовтың "Статистический аспект содержания текста и его формальное представление" атты мақаласындағы ұсынылған критерийлер [Зубов, 1986, 75-94-бб.]. Осы мәселеге қатысты қай ғылыми ізденісті алмасақ та, олардың негізінде мынандай ой-пікір жатқанын аңғаруға болады: ***мәтін ішіндегі мазмұнға қатысты негізгі тірек сөз сияқты ең маңызды бірлік, алдымен, мәтін ішінде жиі қолданыста болуы қажет және, екіншіден, ол бірлік көптеген абзацтардан тиісті өз орнын алулары керек.*** Міне, осы пікір біздің зерттеуімізге арқау болды. Басқа да шет тілдерін зерттеушілермен бірге, біздің ізденістеріміздің негізінде мұндай мәнді **НТС**, көп жағдайда, суреттеу, баяндау және пайымдау типті мәтіндердегі басты кейіпкерді не негізгі нысанды бейнелейтін атау сөздер. Мәтіндегі сөз қолданыстар ішінен ажыратып алынған НТС, мазмұнды ашу процесіне келгенде, микрожағдаяттардың құрастырушылары ретінде басқа да элементтерді шоғырландыратын орталық рөлін атқарады. Мұндай құрастырушы элементтер ретінде – етістік, сын есім, үстеу және кейбір басқа да сөз таптарына қатысты сөздер болып табылады [Новиков, 1981, 53-б.].

Жағдаятқа байланысты **басты референттер** (оларға НТС сәйкес келеді) төменгі дәрежедегі референттермен өзара байланыста болады. Ал мұндай салыстырмалы сипаттағы **"төмен"** дәрежелі референтті бейнелейтін сөздер, зерттеу тәжірибесінде, екінші дәрежелі тірек сөздер деп немесе **қосалқы тірек сөздер** (ҚТС) деп аталуын жөн көрдік [Жубанов, 2002, 147-150-бб.]. Әрине, ҚТС-дер мәтінде қолдану жиілігі жағынан НТС-ге қарағанда төмен сатыда тұрады және аз абзацтарды қамтиды. Ал мәтіндегі НТС мен ҚТС-ден басқа сөздерге, біз **"басқа сөздер"** сөз тіркесінің қысқартылған түрі – "БС" арқылы белгілеуді дұрыс деп таптық.

Тәжірибе арқылы мәтіннен бөлініп алынған НТС және ҚТС арнайы жүргізілетін **"негізгі статикалық мазмұн кестесі"** (НСМК) деп аталатын кесте бойынша көрініс табады.

Қалам иесі, өзінің ой-жоспарындағы мазмұнды құру үшін нақты лексикалық топ (сөздер жиыны) негізінде мәтін ішінде сөздерді "үлестіреді" (таратады) деп жорамалдауға болады. Шындығында, автор сөздерді мәтін бойына таратуды алдын ала ойластырмағанымен, ол ойлау-сөйлеудің ішкі заңдылығынан туындайды. Автор, осы мәтін ішіндегі шағын оқиға туралы айтпақ ойын "таңдап алған" сөздері мен оларды дұрыс үлестірудің негізінде, болуы мүмкін қабылдаушының санасында да бірдей көрініс табу жағын үнемі ескеріп отырады. Автордың нақты сөздерді іздестіре білуі және оларды әртүрлі сөз тіркестерінде қолдануы мәтіннің тақырыбына, яғни онда "кім туралы" және "не туралы" болуына көп байланысы бар.

Біздің түсінігімізше, "тақырып" дегеніміз мәтіннің мағыналық өзегі. Құрылымдық жағынан, ол, негізінен, бас кейіпкер немесе бас нысан мен пікір баяндаушысынан тұрады да, әрі жинақы, әрі абстракты пішіндегі сөздер арқылы берілген мағыналық кешен түрінде көрініс табады. Басқаша айтқанда, тақырып – мәтінде сөз болатын оқиғаны суреттейтін тілдік элементтер мен олардың өзара қатынастарының ең маңызды деген құрастырушыларына жататын сөздер мен сөз тіркестер жиынтығы [Смит, 1980, 336-б.]. "Тақырып", мәтін иесі мен оқырман-

көпшіліктің мәтін бойынан бірдей бөліп алатын және есте қалуға ыңғайлы мағыналық бөлігі. Мәтін бойында мұндай мағынаның инвариантты бөлігінің болуын – бірнеше сынақ ретінде алынған мәтін мағынасын талдау мен жинақтау тәжірибесі жүргізген психолингвистердің де қолдайтыны белгілі болды [Бондаренко, 1978, 80-91-бб.].

Тәжірибесіз, интуитивті жолмен қарастырғанның өзінде, әрбір мәтіннің бір ғана мағыналық тұтастық екенін аңғаруға болады. Себебі оның барлық бөліктерінде "бір ғана мәселе" жайлы ой өрбиді, яғни түгел мәтін бойында біркелкі кейіпкерлер мен нысандарға қатысты жағдаяттар баяндалып отырады және олар оқиға желісінің соңына дейін шектеулі уақыт пен санаулы орынға қатысты суреттеліп отырады. Ал автордың жеке көзқарасы бойынша суреттелген оқиғаны оқырманға (қабылдаушыға) жеткізу, оның жеке мақсаты мен ниетіне байланысты жоғарыда аталған оқиға құрастырушы элементтер үшін олар тең дәрежеде болмауы да мүмкін. Мұндай теңсіздіктің нәтижесінде кейбір кейіпкерлер бір ғана уақиғаны суреттейтін микрожағдаяттардың көбінде қатысуы мүмкін де, ал қалғандары тек кейбір шағын оқиғаның кейбіреуіне ғана қатысты болуы мүмкін. Осы айтылғанның нәтижесінде аталмыш құрастырушы элементтердің мәтін бойындағы атаулары біркелкі дәрежеде болмай, денотаттардың иерахиялық дәрежесіне сәйкес келетін, өзінше бір бөлек иерахиялы атаулар құрылады [Лингвистические ..., 1983, 123-б.; Новиков, 1981, 55-б.].

Ең алдымен, әрбір оқиға желісі нақтылы автордың шығармашылығына тікелей қатысты екенін, сондықтан оның баяндамақшы болған оқиғасы, көбінде, мәтін иесінің өмірлік тәжірибесіне, әлемді қабылдау дәрежесіне сай көрініс тауып, мәтін динамикасына ұйытқы болады [Степанов, 1981, 323-б.]. Осы себепті, автордың жеке өзіне қатысты мәтіндегі атаулар, жиілігінің аз-көптігіне қарамай, **негізі тірек сөз** болып саналады. Сонымен қатар, сөздердің бұл тобына мына мағынадағы атауларын да жатқызуды ұйғардық:

1) бас кейіпкерлер атауы;
2) іс-әрекетке қатысты негізгі нысандар атауы;
3) іс-әрекеттің болған мекен атаулары;
4) іс-әрекетке қатысты мезгіл атаулары.

Әдетте мәтін бойында аталған сөздер тобы жоғары қолданыста кездеседі және көптеген мәтін үзінділерін (абзацтарды) қамтиды. Орыс тіліндегі әдебиеттер тілімен айтсақ, олар қажетті **"изотопты тізбектер", "тақырыпты прогрессиялар", "номинативті тізбек", "рекурентті тізбек", "семантикалық изотоптар", "имплицитті референция", "лексика-тақырыптық сызықтар", "тізбектеу"** және т.б. құрайтыны белгілі [Келемен, 1977, 121-б.; Проблемы ..., 1983, 49-б.].

Мұндай лексикалық бірліктер тізбегінің негізгі ерекшеліктері – мазмұн жағынан тек бір ғана заттық атауға қатысты болуы. Бұл қатарда толық түрдегі сөздік синонимдері, контекстік синонимдер, ассоциативті және есімдіктер арқылы жасалатын алмастырулар және т.б. өзгерістер түрлері орын алуы мүмкін [Леонтьева, 1981, 23-25-бб.].

Жоғарыда аталған төрт түрлі тірек сөздер тобын анықтайтын атаулар денотативтік сәйкестендіру әдісіне қатысты екені белгілі. Мәтін мазмұнын талдауда атау сөз бен сөз тіркестерінің **"маңыздылық дәрежесін"** анықтаудың бірнеше әдіс-тәсілдері бар [Белоногов, Кузнецов, 1983, 135-236-бб.; Зубов, 1985, 176-б.].

Оларды мына топтарға бөлуге болады:
1) анкеттік;
2) құрылымдық;
3) сөздік арқылы;
4) жиілігіне қарай;
5) синтаксистік қызметіне қарай;
6) диалогтар негізінде.

Бұл аталған әдістердің басты кемшілігі – мәтінді біртұтас мағыналық байланыстағы бірлік деп қарастыруында. Мысалы, құрылымдық әдіс арқылы тірек сөздерді бөліп алу, олардың тақырыпқа не болмаса мәтіннің бастапқы сөйлемдеріне ену не енбеу принципін ұстаным етуі. Жиілікті ескеру әдісі бойынша – сөздің мәтіндегі қолдану дәрежесіне қарай, тірек сөздерді ажырату. Мұндай принциптерді ұстам ету – мәтін мағынасынан гөрі, оның пішіндік берілуі қасиеттерін көбірек ескеру деген сөз [Зубов, 1985, 176-б.]. Аталған әдістердің басқа түрлерінің де өздеріне тән кемшіліктері жоқ емес.

Қазақ мәтіндерінен тірек сөздерді бөліп алу әдісіне арнайы тоқталайық. Жоғарыда сөз болғандай, тірек сөздерді мәтін бойынан бөліп алудың басты критерийі – сөздің мәтін ішіндегі абсолютті қолдану жиілігі (алмастырулар ескеріледі) және сол сөз орын алған абзацтар саны.

Біздің зерттеуімізде, қазақ мәтіндерінен тірек сөздерді бөліп алу үшін терминдердің статистикалық тұрақтылығын анықтайтын **маңыздылық коэффициенті** деп аталатын өрнектің өзгертілген түрі пайдаланылды [Жубанов, 2002 150-151-бб.]:

$$\boldsymbol{K}_{маң} = \frac{F \cdot m}{N \cdot n} \quad . \qquad (1),$$

мұндағы F – мәтін ішіндегі сөздің (сөзтұлғаның) абсолютті жиілігі (алмастырулар ескеріледі); m – санаққа ілінген сөзге қатысты абзац саны; N – мәтін ішіндегі сөз қолданыс саны; n – мәтін бойындағы абзацтардың жалпы саны. $\boldsymbol{K}_{маң}$ – сөздің маңыздылық коэффициенті.

Мәтін ішіндегі тірек сөздерді басқа сөздерден ажыратып алу үшін маңыздылық коэффициентінің мәні анықталады және нақты сөздердің жиілігі бойынша олардың сол жақтық және оң жақтық шекараларын формалды түрде анықтайтын $\boldsymbol{K}^1_{маң}$ және $\boldsymbol{K}^2_{маң}$ өрнектері мына пішінде көрініс табады:

$$\frac{\left(\frac{n}{m_1}+1\right)^2}{N \cdot n} \le \boldsymbol{K}^1_{маң} \le 1, \qquad (2);$$

$$\frac{\left(\frac{n}{m_2}+1\right)^2}{N \cdot n} \le \boldsymbol{K}^2_{маң} < \frac{\left(\frac{n}{m_1}+1\right)^2}{N \cdot n}, \qquad (3).$$

Бұл жерде (2) теңсіздігі НТС шекарасын, ал (3) – ҚТС шекарасын анықтау үшін қолданылады. Теңсіздіктердегі m_1 мен m_2 мәндері мәтін ішіндегі абзацтардың жалпы санына (n) қатысты болғандықтан, олардың нақты мәні, әр жағдайға байланысты, тәжірибе жүзінде анықталып, арнайы кестеден алынады. Мысалы,

НТС үшін, яғни $K^1_{маң.}$ қатысты мәтіндегі жалпы абзац саны *100-ге* тең не үлкен болса *(n≥100),* онда $m_1=42$, ал *n* мәні *80* санынан үлкен, бірақ *100*-ден кіші болса $m_1=30$ және т.т.

НТС пен ҚТС үшін m_1, m_2 мәндері толық түрде *1-кесте* мен *2-кестеде* көрініс тапты.

Сонымен, мынадай түсініктегі критерийлерді атауға болады:

1) негізгі тірек сөздерге (НТС) жататын сөздер *1-кестедегі* шарттарды қанағаттандырулары қажет;

1-кесте. Қазақ мәтіндеріндегі сөздерді негізгі тірек сөздерге (НТС) жатқызу критериясы

Рет саны	Мәтіндегі абзац саны: ***n***	Тірек сөз кездескен абзацтар саны: m_1	Мүмкіндікті сөздерді НТС тобына жатқызуға қажетті $K^1_{маң.}$ шекаралық мәндері
1	$n\geq 100$	$m_1=42$	$(\frac{n}{42}+1)^2.\frac{1}{N\cdot n}\leq K^1_{маң.}<1$
2	$81\leq n\leq 99$	$m_1=30$	$(\frac{n}{30}+1)^2.\frac{1}{N\cdot n}\leq K^1_{маң.}<1$
3	$51\leq n\leq 80$	$m_1=17$	$(\frac{n}{17}+1)^2.\frac{1}{N\cdot n}\leq K^1_{маң.}<1$
4	$31\leq n\leq 50$	$m_1=16$	$(\frac{n}{16}+1)^2.\frac{1}{N\cdot n}\leq K^1_{маң.}<1$
5	$15\leq n\leq 30$	$m_1=12$	$(\frac{n}{12}+1)^2.\frac{1}{N\cdot n}\leq K^1_{маң.}<1$
6	$9\leq n\leq 14$	$m_1=5$	$(\frac{n}{5}+1)^2.\frac{1}{N\cdot n}\leq K^1_{маң.}<1$
7	$6\leq n\leq 8$	$m_1=4$	$(\frac{n}{4}+1)^2.\frac{1}{N\cdot n}\leq K^1_{маң.}<1$
8	$n=4;\ n=5$	$m_1=3$	$\frac{9}{N\cdot n}\leq K^1_{маң.}<1$
9	$n=3$	$m_1=2$	$\frac{4}{N\cdot n}\leq K^1_{маң.}<1$
10	$n=1;\ n=2$	$m_1=1$	$\frac{3}{N\cdot n}\leq K^1_{маң.}<1$

2) қосалқы тірек сөздерге (ҚТС) жататын сөзде 2-кестедегі шарттарды қанағаттандырулары қажет;

3) *1*-ші және *2-кестелердің* шарттарына бағынбайтын сөздер тобы **"басқа сөздер"** (БС) тобына жатады.

2-кесте. Қазақ мәтіндеріндегі сөздерді қосалқы тірек сөздерге (ҚТС) жатқызу критериясы

Рет саны	Мәтіндегі абзац саны: n	Тірек сөз кездескен абзацтар саны: m_2	Мүмкіндікті сөздерді ҚТС тобына жатқызуға қажетті $K^2_{маң.}$ шекаралық мәндері
1	2	3	4
1	$n \geq 100$	m_2=80	$(\frac{n}{80}+1)^2.\frac{1}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
2	$81 \leq n \leq 99$	m_2=64	$(\frac{n}{64}+1)^2.\frac{1}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
3	$51 \leq n \leq 80$	m_2=38	$(\frac{n}{38}+1)^2.\frac{1}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
4	$31 \leq n \leq 50$	m_2=36	$(\frac{n}{36}+1)^2.\frac{1}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
5	$15 \leq n \leq 30$	m_2=30	$(\frac{n}{30}+1)^2.\frac{1}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
6	$9 \leq n \leq 14$	m_2=14	$(\frac{n}{14}+1)^2.\frac{1}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
7	$6 \leq n \leq 8$	m_2=8	$(\frac{n}{8}+1)^2.\frac{1}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
8	$n=4; n=5$	m_2=2	$\frac{4}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
9	$n=3$	m_2=1	$\frac{3}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$
10	$n=1; n=2$	m_2=1	$\frac{2}{N \cdot n} \leq K^{2}_{маң.} < K^{1}_{маң.}$

Аталған екі топтағы (НТС және ҚТС) тірек сөздер мағыналық жағынан біркелкі емес. Бұлай тұжырымдаудың негізгі себебі – референттердің заттық қатынастық қасиеттеріне қарай жанды, жансыз нысандар және орынға, уақытқа қатысты сөздер топтары болып ажыратылуы. Тек қана осы сөздер пікір баяндаушымен бірге (предикатпен) ақиқатты түрде бар категориялардың, яғни материя (объективті дүние), қозғалыс, уақыт және кеңістіктің негізгі қарым-қатынасын анықтайды [Лосева, 1980, 15-51-бб.].

ҚОРЫТЫНДЫ

Тіл білімі қандай ғылым саласы деген сауалға жауап ретінде мынаны айтуға болады. Дәл және нақты ғылымдар уақыт көшімен дами отырып, ағашқа ұқсап өседі: кейбір бұтақтары қурап түсіп жатса, кейбіреуі жайқалып желек жаяды және де ағаш діңінің жуандауына байланысты оның тамыры да тереңге – тарих қойнауына кетеді. Тіл білімінің дамуы бұдан өзгеше – бұл кең көгалдағы құлпырған гүлдердің өрнегі іспеттес және де бұл көгал сыйқырға толы: жаңа гүлдер шыққанда ескілері солып қалмайды, өзінің айқын да балауса күйін жоғалтпайды. Міне бұл айтылғандардан қолданбалы тіл білімінің жаңа саласының қазіргі кездегі дамуы бұрын қол жеткен нәтижелер негізінде "жұпар иіске" ие болуда. Гүл түрлерінің, хош иістерінің әртүрлі болуы тәріздес, мәтіндік бірліктердің мән-мағынасы мен мазмұнын лингвистикалық автоматта берілу тәсілдері де әр бағытта дамып келеді. Солардың ішінде мәтін лингвистикасы тұрғысынан қарастыру жолы ғылымның соңғы жетістігі деуге болар лықтай. Осыған қатысты тілді зерттеуге динамикалық көзқараспен қарау, әсіресе, айтылымның туындауы мен әртүрлі түсінікке ие болу сипаты және оны тілдік қарым-қатынас тұрғысынан қарастыру нәтижелері аса құнды ғылыми жетістік болуда.

Мәтіннің құрылымы мен құрастырушыларының ойлау процесіне, ақиқат шындыққа қатыстылығын ескеру арқылы ғана мәтіннің тек өзіне ғана тән қасиеттерін, бірліктерін, мазмұнын ажыратып алуға болады. Осындай қасиеттерге ие болып келетін мәтін бірліктері семантикалық, грамматикалық және логикалық жағынан бір-біріне тәуелді болып келеді.

Мәтіннің мазмұнын – мәтін ішінде кездесетін атаулардың сызықтық тізбегінің оқырманның сана-сезіміндегі бейнеленуі деп те түсіндіруге болады. Ал жалпы таным теориясындағы, тіл қызметіндегі және әртүрлі мәтіндік құрылымдардағы ықтималдықтар теориясының алатын рөлін қысқаша былайша қорытындылауға болады:

- ақиқат әлемдегі барлық байланыс түрлері біркелкі құбылыстар жиынында көрініс табатын объективті сипатқа ие болады;
- тіл, өзін-өзі реттеуші жүйе болуынан, туындау процесі кезінде әрі детерминденген ережелерді, әрі кездейсоқ факторларды қолдану арқылы әрекет етеді;
- мәтін сөйлеу тілі қызметінің нәтижесі ретінде детерминделінген және ықтималды құрастырушылардан тұрады;
- мәтін абзацтары мазмұн "кванты" және "тұрақты" синтаксистік шаблон (үлгі) ретінде ықтималды-алгоритмдік принципте біртұтастық құрайды.

Соңғы айтылғандағыдай, абзацты ең кіші семантика-синтаксистік мәтін бірлігі деп қарастырудың негізінде мәтін абзацтарын олардың функционалды-мағыналық мазмұны бойынша, ал құрылымдық орнына қарай – *бастапқы, медиалды, қорытынды абзацтар* және *заттық-логикалық пішіндегі абзацтар* деп ажыратуға мүмкіндік туады.

Зерттеушілер тарапынан сөйлем түрлерінің қайсыбір синтаксистік және семантикалық модельдерінің көптеген алгоритмдері мен бағдарламаларының құрастырылғаны мәлім. Бірақ олардың барлығында да синтаксистік және лексикалық

бірліктерді таңдауда қатаң түрдегі критерий жоқ. Ал дұрыс мағыналы сөйлемдерді формалды жолмен құру үшін мәтінді туындатудың детерминденген және кездейсоқ процестерін тек қиыстыру мен қисындастыру арқылы іске асатындығын негіз етіп алу керек [Зубов, 1989].

Автордың әрқилы мәтіндерді туындатуы, кейбір жағдайларда, формалды принциптер арқылы іске асатындығы мәлім. Осы айтылғанға қарапайым мысал ретінде мынадай қазақ ертегісінен үзінді келтірейік:

"Ерте-ерте ертеде, ешкі құйрығы келтеде, бір "А" болыпты. "А"-ның үш баласы бар екен. Ең үлкенінің аты "Б", ортаншысының аты "В", ал кенже баласынікі – "С" екен. "С" өте $<\alpha>$, $<\beta>$, $<\gamma>$, *әрі* $<\rho>$ *бала болып өседі. "А" ұйықтап жатып түс көреді, түсінде бір ғажайып <d> көреді, <d> күлсе, аузынан будақ-будақ <e> түседі, жыласа көзінен <z> түседі. Міне, "А" осы <d>-ға ынтық болады. Дереу орнынан тұрысымен "Б" мен "В" -ны шақырып алып, түсін айтып береді...".*

Шындығында, бұл ертек үзіндісіндегі: *"А" – патша, "Б" – Асан, "В" – Үсен, "С" – Хасан.* Сол сияқты $<\alpha>$ *сұлу,* $<\beta>$ *ақылды,* $<\gamma>$ *батыр,* $<\rho>$ *ақ көңілді, <d> –құс, <e> – гүл, <z> – моншақ-моншақ маржан* және т.т.

Енді *"А, Б, В, С"* және $<\alpha, \beta, \gamma, \rho, d, e, z>$ заттық белгі мен сындық белгіні білдіретін формалды бірліктердің мән-мағынасын басқаша өзгертіп, құбылта түссек, осы типті қазақ ертегілерінің де түр-түрін шығаруға болар еді.

Әрбір тілдің өзіндік ішкі функционалды стильдерінің болатыны белгілі. Орыс тілінде жазылған ғылыми әдебиеттерде ондай кішігірім стильді "подъязык" терминімен атайды. Тілдің толық түріне қарағанда "подъязык" шекті лингвистикалық бірліктерден тұрады және өздеріне тән қолданыс аясы болады. Осы бірліктердің әрі ең жиі қолданылғандарын, әрі мазмұндылық жағынан маңыздыларын және оларға тән келетін мәнмәтіндерді (контекстерді) іріктеп алуға болады. Міне, осындай іріктеліп алынған тілдік элементтер негізінде табиғи тілдік жүйеге жақын жатқан және шекті мазмұндық тақырыптағы тілдік модельдерді туындатуға болатындығы ғылыми жолмен дәлелденіп отыр.

Осы айтылғандарға қатысты немістің ірі лингвисті В.Гумбольдт былай дейді: "... Язык имеет некий предел законченности организации, после достижения которого уже не подвергаются никаким изменениям ни его органическое строение, ни его структура... Еще не было обнаружено ни одного языка, находящегося ниже предельной границы сложившегося грамматического строения. ..." [Звегинцев, 1964-1965 бойынша].

Сонымен, қазақ тіліне қатысты белгілі құжаттық мәтіндерді талдау мен жинақтау әдістері арқылы жүргізілетін ғылыми зерттеулер – бүгінгі таңның уақыт күттірмейтін маңызды ісі.

ӘДЕБИЕТ

1. *Агроскина С.Н.* Структура абзаца в художественной прозе: Автореф. дис. канд., Л., 1972. 25 с.
2. *Адамец П.* Образование предложений из пропозиций. Прага, 1978.
3. *Адмони В.Г.* Введение в синтаксис современного немецкого языка. М.: Изд. лит. на иностр. языках, 1955. 392 с.
4. *Адмони В.Г.* Качественный и количественный анализ грамматических явлении // Теоретические проблемы современного советского языкознания. М., 1964. С. 56-68.
5. *Адмони В.Г.* Синтаксическая семантика – это семантика, синтаксических структур // Проблемы синтаксической семантики. М., 1976. С. 3-8.
6. *Алейник А.* От интеграла до рубля // Молодой коммунист. 1974. № 5. С. 50-57.
7. *Алексеев П.М.* Квантитативные аспекты речевой деятельности // Языковая норма и статистика. М., 1977. С. 43-58.
8. *Алексеев П.М., Манасян Н.С.* Лингвистические распределения (элементы количественного анализа текста). Л., 1985.
9. Анализ речевых сигналов человеком /Отв. ред. Г.В. Гершуни. М., 1971.
10. *Андреев Н.Д.* Вероятностные дифференциальные признаки и аппроксимационный анализ синтаксиса // Статистико-комбинаторное моделирование языков. М.-Л., 1965. С. 340-346.
11. *Андреев Н.Д., Зиндер Л.Р.* Основные проблемы прикладной лингвистики. ВЯ, 1959. № 4.
12. *Анисимов С.Ф.* Человек и машина. М., 1959.
13. *Анохин П.К.* Опережающее отражение действительности // Вопросы философии. 1962. № 7. С. 14-25.
14. *Апресян Ю. Д.* 1) Синтаксис и семантика в синтаксическом описании // Единицы разных уровней грамматического строя языка и их взаимодействие. М., 1969; 2) Лексическая семантика. М., 1974.
15. *Арапов М.В.* Квантитативная лингвистика. М., 1978.
16. *Арутюнова Н.Д.* Понятие пропозиции в логике и лингвистике // Изв. АН СССР. Серия лит. и языка. 1976. № 1. Т. 35;
17. *Арутюнова Н.Д.* Предложение и его смысл. М.: Наука, 1976. 383 с.
18. *Балли Ш.* Общая лингвистика и вопросы французского языка / Пер. с фр. М., 1955. 416 с.
19. *Балонов Л.Я., Деглин В.Л.* Слух и речь доминантного и не доминантного полушарий. Л., 1976.
20. *Бару А.В.* Слуховые центры и опознание звуковых сигналов. Л., 1978.
21. *Бектаев К.Б., Зубов А.В., Ковалевич Е.Ф.* и др. К исследованию законов распределения лингвистических единиц // Статистика текста. Минск, 1969. С. 131-162;
22. *Белоногов Г.Г., Кузнецов Б.А.* Языковые средства автоматизированных информационных систем. М.: Наука, 1983. 287 с.
23. *Белый Андрей.* 1) Мастерство Гоголя. М. – Л., 1934; 2) Ритм как диалектика и "Медный всадник". М., 1929.
24. *Бернштейн Н.А.* Очерки по физиологии движений и физиологии активности. М.: Медгиз, 1966.
25. *Блумфильд Л.* Язык. М.: Прогресс, 1968. 606 с.
26. *Богданов В.В.* 1) Актанты и сирконстанты // Проблемы членов предложения в индоевропейских языках. Грозный, 1978; 2) Семантико-синтаксическая организация предложения. Л., 1977; 3) Механизм семантической деривации // Семантико-синтаксическая организация предложения и текста. Грозный, 1980; 4) Речевое общение: Прагматические и семантические аспекты. Л., 1990. 5) Текст и текстовые общение. СПб., 1993.
27. *Богданов В.В.* Статистические концепции языка и речи // Статистика речи и автоматический анализ текста. 1972. Л., 1973. С. 9-19.
28. *Богданов В.В.* Моделирование семантики предложения. В кн.: Прикладное языкознание. Учебник, отв. ред А.С.Герд. Изд. Санкт-Петер. универ, 1996. С. 161-200.
29. *Бондаренко В.Н.* Отрицание как логико-грамматическая категория. М., 1983.
30. *Бондаренко Г.В., Шрейдер Ю. А.* Текст-смысл-ситуация (к постановке проблемы). В кн.: Вопросы информационной теории и практики. № 36. М.: ВИНИТИ, 1978. С. 80-91.

31. *Бондарко А.В.* Принципы функциональной грамматики и вопросы аспектологии. Л.: Наука, 1983. 208 с.
32. *Бондарко А.В.* Фонетические аспекты прикладной лингвистики. В кн.: Прикладное языкознание. Учебник., /Отв. ред. А.С.Герд. Изд. Санкт-Петер. универ, 1996. С. 113-141.
33. *Бондарко Л.В.* 1) Звуковой строй современного русского языка. М., 1977; 2) Фонетические описание языка и фонологическое описание речи. Л., 1981.
34. *Бондарко Л.В., Вербицкая Л. А., Гордина М.В.* Основы общей фонетики. СПб., 1991.
35. *Бондарко Л.В., Загоруйко Н. Г., Кожевников В.А., Молчанов А.П., Чистович Л.А.* Модель восприятия речи человеком. Новосибирск, 1968.
36. *Борель Эмиль.* Вероятность и достоверность. М.: Наука, 1964. 119 с.
37. *Борисов В.В.* Аббревиация и акронимия. М., 1972.
38. *Борода М.Г.* О характере распределения информационных единиц малой частоты в художественных текстах // Семиотика и информатика. Вып. 9. М., 1977.
39. *Брандес О.П.* Прагматика языка как переводческая проблема. Тетради переводчика. М., 1979, Вып. 16. С. 65-71. Проблема текстуальной лингвистики. Киев: КГУ, 1983. 175 с.
40. *Бриллюэн Л.* Наука и теория информации. М., 1969.
41. *Булыгина Т.В., Селиверстова О.Н., Елисеева А.Г.* и др. Семантические типы предикатов. М., 1982.
42. *Буторов В.Д.* Приближенные методы анализа языка в прикладной лингвистике // Структурная и прикладная лингвистика. Вып. 2. Л., 1983. С. 131-137.
43. *Буторов В.Д.* Моделирование синтаксиса естественного языка. В кн.: Прикладное языкознание. Учебник /Отв. ред. А.С.Герд. Изд. Санкт-Петер. универ, 1996. С. 142-160.
44. Бюллетень фонетического фонда русского языка / Под ред. К. Саплока и Л. Бондарко. Вып. 1. Л.: Бохум, 1988; Вып. 2. 1989; Вып 3. 1990; Вып. 4. 1992.
45. *Ван Валин Р., Фоли У.* Референциально-ролевая грамматика //Новое в зарубежной лингвистике. Вып.XI. Современные синтаксические теории в американской лингвистике. М., 1982.
46. *Васильев Л.М.* Речевой смысл предложения как результат взаимодействия языкового и неязыкового знания. В сб.: Предложение и текст: семантика, прагматика и синтаксис. Л.: 1988. С. 5-10.
47. *Винокур Г.О.* Избранные работы по русскому языку. М.: Учпедгиз, 1959. 492 с.
48. *Винокур Т.Г.* Говорящий и слушающий. М., 1993.
49. *Витгенштейн Л.* Логико-философский трактат. М., 1958.
50. *Вихляев П.А.* Очерк теоретической статистики. М., 1928.
51. *Волошин Е.П.* Аббревиатуры в лексической системе английского языка: АКД. М., 1967.
52. *Воронин С.В.* Фоносемантические идеи в зарубежном языкознании. Л., 1990.
53. Восприятие: Механизмы и модели / Пер. с англ.; Под. ред. П. Ю. Алексеенко. М., 1974.
54. *Гак В.Г.* 1) К проблеме синтаксической семантики (семантическая интерпретация "глубинных" и "поверхностных" структур) // Инвариантные синтаксические значения и структура предложения. М., 1969; 2) Высказывание ситуация // Проблемы структурной лингвистики. 1972. М., 1973.
55. *Гак В.Г.* О семантическом инварианте предложения. – В кн.: Проблемы синтаксической семантики. –М.: МГПИИЯ им. М.Тореза, 1976. С. 22-26.
56. *Гальперин И.Р.* О понятии текст // Вопросы языкознания. 1974. № 6.
57. *Гальперин И.Р.* Текст как объект лингвистического исследования. М.: Наука, 1981. 139 с.
58. *Гендина Н.И.* Выявление аспектов содержания документов с помощью формальных текстовых признаков // НТИ Сер. 2. 1979. № 4. С. 7-12. *Скороходько Э.Ф.* Семантические сети и автоматическая обработка текста. Киев: Наукова думка, 1983.
59. *Герд А.С.* Логико-понятийная структура научного текста и его семантика. В кн.: Вопросы семантики. Калининград, 1978. – С. 55-58.
60. *Герд А.С.* Основы научно-технической лексикографии: Как работать над терминологическим словарем. Л., 1986. 73 с.
61. *Герд А.С.* Научно-техническая лексикография. В кн.: Прикладное языкознание. Учебник. /Отв. ред. А.С.Герд. Изд. Санкт-Петер. универ, 1996. С. 287-307.
62. *Гладкий А.В.* 1) Синтаксические структуры естественного языка в автоматизированных системах общения. М., 1985; 2) Формальные грамматики и языки. М., 1973.
63. *Гладкий А.В., Мельчук И.А.* Элементы математической лингвистики. М., 1969.
64. *Глушков В.М.* Мышление и кибернетика. В сб.: Диалектика в науках и неживой природе. М., 1964.
65. *Головин Б.Н.* Язык и статистика. М., 1971.

66. *Грайс Г.П.* Логика и речевое общение //Новое в зарубежной лингвистике. Вып. 16. Лингвистическая прагматика. М., 1985.
67. *Гринев С.В.* Введение в терминологическую лексикографию: Учеб. пособие. М.,1986. 103 с.
68. *Гришина О.Н.* Проблемы контекстно-вариативного членения текста в стиле языка художественной и научной прозы // Функциональные стили и преподавание иностранных языков. М.: Наука, 1982. С. 52-68.
69. *Гузман А.* Интегральные работы / Пер. с англ. Вып. 2. М., 1975. 528 с.
70. *Даниленко В.П., Волкова И.Н., Морозова Л.А., Новиков Н.Р.* Лингвистический аспект стандартизации терминологии. М., 1993.
71. *Деркач М.Ф., Гумецкий Р.Я., Гура Б.М., Чабан М.Е.* Динамические спектры речевых сигналов. Львов, 1983.
72. Детский голос / Под ред. В.Н. Шацкой. М. 1970.
73. *Джубанов А.Х.* Квантитативная структура казахского текста (опыт лингвистического анализа на ЭВМ). Алма-Ата, 1987. С. 147.
74. *Долгова О.В.* Синтаксис как наука о построении речи. М.: Высшая школа, 1980.
75. *Ершов А.П.* К методологии построения диалоговых систем: феномен деловой прозы. Предпринт. 156. Новосибирск: ВЦСОАН СССР, 1979. 24 с.
76. *Жинкин Н.И.* Сенсорная абстракция // Проблемы общей, возрастной и педагогической психологии. М.: Педагогика, 1978.
77. *Жубанов А.К.* Основные принципы формализации содержания казахского текста. Алматы, 2002. 250 с.
78. *Жуков Н.И.* Философские основы кибернетики. Минск: БГУ, 1973.
79. *Журавлев А.П.* Фонетические значение. Л., 1974.
80. *Закиев М.З.* Актуальные проблемы в сложных предложениях в тюркских языках. Баку, 1983.
81. *Зарубина Н.Д.* О психологическом обосновании приемлемости предложения и сверхфразового единства в качестве единиц обучения иностранному языку // Психология грамматики. М.: МГУ, 1968.
82. *Зарубина Н.Д.* Текст: лингвистический и методический аспекты. М.: Русский язык, 1981. 113 с.
83. *Звегинцев В.А.* Предложение и его отношение к языку и речи. М.: МГУ, 1976. –307с.
84. *Звегинцев В.А.* Язык и лингвистическая теория. –М.: МГУ, 1973.
85. *Звегинцев В.А.* История языкознания XIX–XX веков в очерках и извлечениях. Тт. 1,2. Изд. 3-е. М.: Просвещение, 1964, 1965.
86. *Зиндер Л.Р.* Общая фонетика. М., 1979.132 с.
87. *Зиндер Л.Р., Строева Т. В.* К вопросу о применении статистики в языкознании // Вопр. языкозн. 1968. № 6. С. 120-123.
88. *Зиновьев А.А, Ревзин И.И.* Лексическая модель как средство научного познания. Вопросы философии. 1960, № 1.
89. *Златоустова Л.В., Потапова Р. К., Грунин-Донской В. Н.* Общая и прикладная фонетика. М., 1986.
90. *Зубов А.В.* Вероятностно-алгоритмическая модель порождения текста (семантико-синтаксический аспект). Минск: 1985. 399 с. (докт. дис. филолог. наук).
91. *Зубов А.В.* Проблемы порождения текста. Часть 1. Теория и алгоритмы. Минск, 1987. 77 с.
92. *Зубов А.В.* Проблемы порождения текста. Часть II. Вероятностно-алгоритмическая модель порождения текста. Минск, 1989. 63 с.
93. *Зубов А.В.* Статистической аспект содержания текста и его формальное представление. В кн: Квантитативная лингвистика и автоматический анализ текстов. Тарту, 1986. С 75-94.
94. *Катенина Т.Е., Рудой В.И.* Лингвистическое знание в Древней Индии // История лингвистических учений: Древний мир. Л., 1980.
95. *Катц Дж.* Семантическая теория // Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика. М., 1981.
96. *Квитко И.С., Лейчик В.М., Кабанцев Г.Г.* Терминоведческие проблемы редактирования. Львов, 1986. 151 с.
97. *Кейтер Дж.* Компьютеры – синтезаторы речи. М., 1985.
98. *Келемен Я.* Текст и значение. В сб.: Семиотика и художественное творчество. М.: Наука, 1977. С. 104-124.
99. *Кибрик А.Е.* Предикатно-аргументные отношения в семантически эргативных языках. Изв. АН СССР. Серия лит. и языка. Т. 39. 1980, № 4. С. 324-335.

100. *Клещев А.С.* Представление знаний. Методология, формализмы, организация вычислений и программная поддержка // Прикладная информатика. Вып. I. М.: Финансы и статистика, 1983.
101. *Клоусон ДЖ.* Лексико-статистическая оценка алтайской теории // ВЯ, 1969. № 5, С.22-41.
102. *Кобрин Р.Ю.* Лингвостатический анализ терминологических систем // Вычислительная лингвистика. М., 1976. С. 144-158.
103. *Кобрина М.А., Малаховский Л.В.* Английская пунктуация. М., 1961.
104. *Колшанский Г.В.* Контекстная семантика. М.: Наука, 1980.
105. *Колшанский Г.В.* О языковом механизме порождения текста // Вопросы языкознания. 1983. № 3.
106. *Кондратов А.М.* Звуки и знаки. М., 1996; Изд. 2-е, М., 1978.
107. *Кузнецов И.П.* Некоторые механизмы порождения текстов описания // НТИ. Сер. 2. 1975. № 9 С. 10-15.
108. *Купина Н. А.* Опыт системно-синтаксического анализа семантики связного текста. В кн.: Семантика и структура предложения. Лексическая и синтаксическая семантика. Уфа: Изд. Башкирского Гос. унив., 1978. С. 137-141.
109. *Қордабаев Т., Қалиев Ғ.* Жалпы тіл білімі. Алматы: Дәуір, 2004. 240 б.
110. *Лакофф Дж.* О порождающей семантике // Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика. М., 1981.
111. *Лейчик В.М., Шелов С.Д.* Лингвистические проблемы терминологии и научно-технический перевод: В 2 ч. Ч. 1. М., 1989. 40 с; Ч. 2. 1990. 78 с.
112. *Лейчик В.М.* Прикладное терминоведение и его направления. В кн.: Прикладное языкознание. Учебник. /Отв. ред. А.С.Герд. Изд. Санкт-Петер. универ, 1996. С. 276-286.
113. *Леонтьев А.А.* Психолингвистические единицы и порождение речевого высказывания. М.: Наука, 1969. 307 с.
114. *Леонтьев А.Н.* Деятельность. Сознание. Личность. М.: Изд-во полит. литер., 1977.
115. *Леонтьева Н.Н.* Семантика связного текста и единицы информационного анализа. НТИ. Сер. 2. 1981, № 1. С. 21-29.
116. Лингвистика на исходе XX века: итоги и перспективы: В 2 т. М., 1995.
117. Лингвистическая концепция Терминологического банка данных Машинного фонда русского языка: (Проект) / Под. ред. проф. А.С. Герда. М., 1989. 104 с.
118. Лингвистические вопросы алгоритмической обработки сообщений. М.: Наука, 1983. 246 с.
119. Логопедия / Под. ред. Л.С. Волковой. М., 1989.
120. *Лосева Л.М.* Как строится текст. Пособие для учителей. М.: Просвещение, 1980. 94 с.
121. *Лоскутов В.И.* Автоматизированные системы управления. М., 1972.
122. *Лукьяненков К.Ф.* Использование схем Пуассона и Гаусса в исследовании распределения лингвистических единиц текста // Вопросы лингвостатистики и автоматизации лингвистических работ. М., 1970. Вып. 2. С. 3-14.
123. *Макколи Дж.* О месте семантики в грамматике языка // Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика. М., 1981.
124. *Малаховский Л.В.* О способах обработки знаков препинания при автоматическом переводе с английского языка на русский. Ст РААТ, 1971.
125. *Малитков А.С., Сольский Д.И.* (ред.) Руководство по делопроизводству. М., 1972.
126. *Мартыненко Г.Я.* Методы статистического моделирования в языкознании. В кн.: Прикладное языкознание. Учебник. /Отв. ред. А.С.Герд. Изд. Санкт-Петер. универ, 1996. С. 201-223.
127. *Марусенко М.А.* Коэффициент статистической устойчивости как критерий отбора учебного материала // Материалы семинара "Статистическая оптимизация преподавания языков и инженерная лингвистика". Чимкент, 1980. С. 253-254.
128. *Марчук Ю.Н.* Проблемы машинного перевода. М.: Наука, 1983.
129. *Маслов Б.А.* Проблемы лингвистического анализа связного текста (межфразовый уровень). Учеб. пособие к спецкурсу. – Таллин, 1975. 104 с.
130. *Медникова Э.М., Шеляховская Л.А.* К типологии аббревиатур. –В кн. Вопросы оптимализации естественных коммуникативных систем. М., 1971.
131. *Мельчук И.А.* Автоматический синтаксический анализ. Новосибирск, 1964.
132. Механизмы речообразование и восприятие сложных звуков / Отв. ред. Г.В. Гершуни. М.; Л., 1966.
133. *Мильчин А.Э.* Методика редактирования текста. М.: Книга, 1980. 318 с.
134. *Морозов В.П.* Биофизические основы вокальной речи. М., 1977.

135. *Москальская О. И.* Грамматика текста. М.: Высшая школа, 1981. 183 с.
136. *Невежина З.В.* Структурно-семантическая организация сверхфразового единства в современном английском языке: Автореф. дис. ... канд. филол. наук. Киев, 1971. 22 с.
137. *Недбаева Г.И.* К вопросу о средствах выражения связи между абзацами // Структура предложения и абзац. Ростов на Дону, 1974.
138. *Нелюбин Л.Л.* Перевод и прикладная лингвистика. М: Высшая школа, 1983. 207 с.
139. *Нелюбин Л.Л.* Информационно-статистические и инженерно-лингвистические особенности языка и тескт в условиях спецкоммуникации: АДД. М., 1975.
140. *Нелюбин Л.Л.* Перевод и прикладная лингвистика. М., 1983.
141. *Нелюбин Л.Л.* Порождающие на различных уровнях. В сб.: Проблемы моделирования языка. Т. 3.2. Уч. зап. ТГУ, № 228. Тарту, 1969.
142. *Нелюбин Л.Л.* Стилистика современного английского языка. М., 1976.
143. *Нечаева О.А.* Функционально-смысловые типы речи (описание, повествование, рассуждение). Улан-Удэ, 1974.
144. *Новиков А.И.* Семантика текста и ее формализация. М.: Наука, 1983. 216 с.
145. *Новиков Э.А., Чистякова Г.Д.* К вопросу о теме и денотате // Известия АН СССР. Серия лит. и языка. Т 40, № 1, М.: 1981. С. 48-56.
146. *Новиков Э.А., Егоров В.С.* Информация и исследователь. Л.: Наука, 1974.
147. *Новиков Э.А., Чистякова Г.Д.* К вопросу о теме и денотате // Изв. АН СССР. Серия лит. и языка. Т 40. № 1. М., 1981. С. 48-56.
148. *Обжелян Н.К., Трунин–Донской В.Н.* Речевое общение и системах "человек – ЭВМ". Кишинев, 1985.
149. Общее языкознание. Формы существования, функции, история языка. М.: Наука, 1970.
150. *Орлов Ю.К.* О статистической структуре сообщений, оптимальных для человеческого восприятия // Научно-техническая информация. Сер. 2. (М.). 1970. № 2. С. 11-16.
151. *Откупщикова М.И.* Моделирование языка // Прикладное языкознание. Учебник / Отв. ред. А.С.Герд. Изд. Санкт-Петер. универ. 1996. С. 100-112.
152. *Падучева Е.В.* 1) Высказывание и его соотнесенность с действительностью. М., 1985; 2) О семантике синтаксиса. М., 1974.
153. *Падучева Е.В.* О способах представления синтаксической структуры предложения // Вопр. языкозн. 1964. № 2.
154. *Панкрац Ю.Г.* Пропозициональные структуры и их роль в формировании языковых единиц разных уровней. Минск, Москва, 1992.
155. *Перекальская Т.К.* Абзац как основная единица научного текста // Вопросы методики и лингвистики. М.: МГУ, 1980.
156. *Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А.* Математическая лингвистика. М., 1977.
157. *Пиотровский Р.Г., Шингарева Е.А.* Семантико-семиотическое сопоставление естественного языка и объектно-признакового информационного языка. В кн.: Романское и германское языкознание. Вып. I. Минск: Изд. МГПИИЯ, 1979. С. 179-191.
158. *Плотников Б.А.* Основы семасиологии. Минск: Вышэйшая школа, 1984. 223 с.
159. *Поспелов Д.А.* Профессионально и проблемно ориентированные интеллектуальное системы // Вопр. философии. 1979. ғ 2. С. 57-59.
160. Прикладное языкознание. Учебник /Отв. ред. А.С.Герд. Изд. Санкт-Петер. универ, 1996.
161. Проблемы текстуальной лингвистики. Киев: КГУ, 1983.
162. Прогноз в речевой деятельности. М.: Наука, 1974.
163. *Разинкина Н.М.* О понятии стереотипа в языке научной литературы // Проблемы повышения эффективности и качество обучения иностранным языкам научных работников. Тезисы докладов. Москва-Цхалтубо, 1978.
164. Разработка стандартов на термины и определения: Рекомендации. Р. С. 50-603-1-89. М., 1990. 54 с.
165. *Рамишвили Г.С., Чикоидзе Г.Б.* Криминалистическое исследование фонограмм речи и идентификация личности говорящего. Тбилиси, 1991.
166. Распознавание слуховых образов / Под. ред. Н. Г. Загоруйко и Г. Я. Волошина. Новосибирск, 1970.
167. *Рассел Б.* Человеческое познание, его сфера и границы. Перевод с англ. М., 1957.
168. *Рахманин Л.В.* Стилистика деловой речи и редактирование служебных документов. М., 1973.
169. *Ревзин И.И.* Модели языка. М., 1962.

170. *Реферовская Е.А.* Коммуникативная структура текста. Л., 1983. 207 с.
171. *Реферовская Е.А.* Коммуникативная структура текста. Л.: Наука, 1989. - 168 с.
172. Речь: Артикуляция и восприятие / Под. ред. В. А. Кожевникова и Л. А. Чистович. М. – Л., 1965.
173. *Романовский В.И.* Применение математической статистики в опытном деле. М.-Л., 1947. 247 с.
174. Русская грамматика. Синтаксис. Т. 2. М., 1980.
175. Русское сценические произношение / Под ред. С. М. Кузминой. М.,1986.
176. *Рылова Т.Н.* Принципы упорядочения и группировка значимых слов текста для совершенствования критерия семантической связи между предложениями. Научно-техническая информация. 1973. № 4. С. 34-38.
177. *Салмина Н.Г.* Знак и символ в обучении. М.: МГУ 1988. 288 с.
178. *Сапожков М.А.* 1) Речевой сигнал в кибернетике и связи. М., 1963; 2) Электроакустика. М., 1978.
179. *Сачков Ю.В.* Введение в вероятностный мир. М., 1971.
180. *Севбо И.П.* Графическое представление синтаксических структур и стилистическая диагностика. Киев, 1981.
181. *Серкова Н.И.* Семантика и метасемантика красной строки в художественной прозе. В кн.: Актуальные вопросы лингвистического и методологического изучения художественного текста. Владимир: 1981. С. 108-120.
182. *Серкова Н.И.* Семиотика абзаца в научной и художественной литературе // Семантическое вопросы микро- и макросинтаксис. Хабаровск, 1980.
183. *Сильдмэя И.* О когитологии // Уч. зап. Тартуск. Гос. унив. Вып. 594. Диалоговые системы и представление знаний. Труды по искусственному интеллекту. IV, Тарту, 1981.
184. *Сильман Т.И.* Структура абзаца и принципы его развертывания в художественном тексте // Теоретические проблемы синтаксиса современных индо-европейских языков. Л.: Наука, 1975.
185. *Сильницкий Г.С.* Семантические типы ситуаций и семантические типы глаголов // Проблемы структурной лингвистики. 1972. М.: Наука, 1973. С. 373-392.
186. *Сиськов В.И.* Корреляционный анализ в экономических исследованиях. М., 1975.
187. *Скалозуб Л.Г.* Динамика звукообразования (по данным кинорентгено-графирования). Киев, 1979.
188. *Смирницкий А.И.* Лексикология английского языка. М., 1956.
189. *Смит ДЖ. Б.* Тематическая структура и тематическая сложность. - В кн.: Новое в зарубежной лингвистике. Вып. IX. М.: Прогресс, 1980. С. 333-355.
190. *Солнцев В.М.* Язык как системно-структурное образование. М.: Наука, 1977. 342 с.
191. *Сорокин Ю.А., Тарасов Е.Ф., Шахнорович А.М.* Теоретические и прикладные проблемы речевого общения. М.: Наука, 1979.
192. *Степанов Ю.С.* В поисках прагматики (проблемы субъекта)//Изд. АН СССР. Серия лит. и языка. Т. 40. 1981. № 4, - 325 с.
193. *Страхова В.С.* Внешние средства организации текста. В кн.: Лингвистика текста. Сборник научных трудов МГПИИЯ им. М. Тореза. Вып. 141. М.: 1971. С. 15-159.
194. *Супрун А.Е.* Лекции по языковедению. Мн.: БТУ, 1978. 144 с.
195. *Сусов И.П.* Семантическая структура предложения. Тула, 1973.
196. *Теньер Л.* Основы структурного синтаксиса. М., 1988.
197. *Тер-Минасова С.Г.* Словосочетание в научно-лингвистическом и дидактическом аспекте. М.: Высшая школа, 1981.
198. *Траугott Н.Н., Кайданова С.И.* Нарушение слуха при сенсорной алалии и афазии. Л., 1975.
199. *Тулдава Ю.А.* Проблемы и методы квантитативно-системного анализа лексики. Таллин, 1987.
200. *Уилсон А., Уилсон М.* Информация, вычислительные машины и проектирование систем. М., 1968. 230 с.
201. Уровни языка в речевой деятельности: К проблеме лингвистического обеспечения автоматического распознавания речи / Под ред. Л.В. Бондарко. Л., 1986.
202. *Ф. де Соссюр.* Курс общей лингвистики. Перевод с франц. М., 1933.
203. *Файгенберг И.М. Иванников В.А.* Вероятное прогнозирование и преднастройка к движениям. М.: МГУ, 1978.
204. *Фант Г.* Акустическая теория речообразование. М., 1964.
205. *Феллер В.* Введение в теорию вероятностей и ее приложения. М., 1967. Т.1. 498 с.

206. Физиология речи: Восприятие речи человеком. Л., 1976.
207. Физиология сенсорных систем. Л., 1972.
208. *Филлмор Ч.* 1) Дело в падеже // Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика. М., 1981 а; 2)Дела о падеже открывается вновь // Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика. М., 1981б.
209. *Фитиалов С.Я.* Формальные грамматики. Л., 1984.
210. *Фланаган Дж.* Анализ, синтез и восприятие речи. М., 1968.
211. *Фрумкина Р.М.* Роль статистических методов в современных лингвистических исследованиях // Математическая лингвистика. М., 1973. С. 156-183.
212. *Фрумкина Р.М.* Вероятность элементов текста и речевое поведение. М., 1971. 168 с.
213. *Фрумкина Р.М.* О законах распределения слов и классов слов // Структурно-типологические исследования. М., 1962. С. 124-133.
214. ФЭС. 1983.
215. *Хайтун С.Д.* Наукометрия: Состояние и перспективы. М., 1983.
216. *Хастингс Н., Пикок Дж.* Справочник по статистическим распределениям / Пер. с англ. М., 1980.
217. *Хомский Н.* Аспекты теории синтаксиса. М., 1972.
218. *Цейтин Г.С.* К вопросу о постарении математических моделей языка // Тез. совещ. по математ. лингвистике 15-21/IV 1959. Л., 1959. С. 45.
219. *Цемель Г.И.* Опознавание речевых сигналов. М., 1971.
220. *Чан Нгок Тхем.* Абзац как грамматико-семантическая сущность. В сб.: Предложение и текст: семантика, прагматика и синтаксис. Л.: Русский язык, 1977. 935 с.
221. *Чахоян Л.П.* Речемыслительные процессы и значение высказывания. В сб.: Предложения и текст: семантика, прагматика и синтаксис. Л.: ЛГУ. 1988. С. 10-15.
222. *Чейф У.* Значения и структура языка. М., 1975.
223. *Чесноков П.В.* Об отношении между речевым и мыслительным процессами с точки зрения единства языка и мышления. В кн.: Синтаксическая семантика и прагматика. Калинин: КГУ, 1982. С. 38-47.
224. *Чупров А.А.* Очерки по теории статистики. СПб., 1910.
225. *Шайкевич А.Я.* Дистрибутивно-статистический анализ текста: Автореф. докт. дис. Л., 1982.
226. *Шаумян С.К.* Операционные определения и их пременение в фонологии. "Применение логики в науке и технике". М., 1960.
227. *Шенк Р.* Обработка концептуальной информации. М., 1980.
228. *Ширяев А.Ф.* Синхронный перевод. М.: Воениздат, 1979.
229. *Школьник Л.С.* Экспериментальное изучение логико-мотивоционной структуры прагматически ориентированного текста // Тезисы VI Всесоюзного симпозиума по психолингвистике и теории коммуникации. Москва, 17 апреля, 1978 г. М.: Наука, 1978.
230. *Шрейдер Ю.А.* Равенство, сходство, порядок. М., 1971.
231. *Шрейдер Ю.А., Шаров А. А.* Системы и модели. М., 1982.
232. *Шрейдер Ю.А.* Текст, автор, семантика и информатика. Вып. 7. –М.: ВИНИТИ, 1976. С. 153-169.
233. *Щерба Л.В.* Языковая система и речевая деятельность. Л.: Наука, 1974.
234. Экспериментально-фонетический анализ речи: Проблемы и методы. Вып. 1. Л., 1984; Вып. 2. 1989.
235. *Яблонский А.И.* Стохастические модели научной деятельности // Системные исследования: Ежегодник. 1975. М., 1975. С. 5-42.
236. *Якобсон Р.* Шифтеры, глагольные категории и русский глагол // Принципы типологического анализа языков различного строя. М., 1972.
237. *Ястремский Б.С.* Некоторые вопросы математической статистики. М., 1960.
238. *Cook W.* Case Grammar: Deveiopment of the Matrix Model (1970- 1978). Washington, 1979.
239. *Herdan G.* Quantitative Linguistics. London, 1964.
240. *Levy P.* Calcul de probabilites. Pt. 2, chap. 6. Gauthier; Villars, 1925.
241. *Lexis W.* Abhandlungen Zur Theorie der Bevolkerungs und Moralstatistik. Vena, 1903.
242. *Zimek R.* Semanticka vystavba vety. Praha, 1980.

КОМПЬЮТЕРЛІК ЛИНГВИСТИКАҒА КІРІСПЕ



Оқу құралында басты назар филология саласында электрондық есептеу машиналарын қолданудың гуманитарлық аспектілеріне аударылады. Еңбекті филологиялық пәндерді оқытуда және жоғарғы оқу орындарындағы әртүрлі арнайы курстар бойынша дәріс беруде пайдалануға болады.

Қолданбалы лингвистиканың дәстүрлі және жаңа салалары: жасанды интеллект, машиналық аударма, ақпарат іздестіру жүйелері, қазақ тілінің компьютерлік қорын құрастыру жайттары және т.б. қарастырылған.

- АЛҒЫ СӨЗ
- КІРІСПЕ
- "АДАМ–КОМПЬЮТЕР–АДАМ" ЖҮЙЕСІНДЕГІ ТАБИҒИ ТІЛ ҚЫЗМЕТІНІҢ ЕРЕКШЕЛІКТЕРІ ЖАСАНДЫ ИНТЕЛЛЕКТ ЖҮЙЕСІН ҚҰРАСТЫРУҒА ҚАТЫСТЫ МӘСЕЛЕЛЕР ЖАЙЫНДА
- ОҚУ ЛЕКСИКОГРАФИЯСЫНЫҢ ҒЫЛЫМИ ПӘН РЕТІНДЕГІ ЕРЕКШЕЛІГІ ҒЫЛЫМИ-ТЕХНИКАЛЫҚ ЛЕКСИКОГРАФИЯ ЖӘНЕ АРНАУЛЫ СӨЗДІКТЕР ТҮРЛЕРІ
- ЛЕКСИКОГРАФИЯ САЛАСЫНДАҒЫ АВТОМАТТАНДЫРУ ТӘСІЛІ
- РЕФЕРАТ ҚҰРАСТЫРУДЫҢ ӘДІС-ТӘСІЛДЕРІ
- АҚПАРАТ ІЗДЕСТІРУ ЖҮЙЕСІ
- АВТОМАТТАНҒАН ОҚУ ЖҮЙЕСІНІҢ НЕГІЗГІ ҰСТАНЫМДАРЫ
- АВТОМАТТЫ (МАШИНАЛЫҚ) АУДАРМА
- БАСПА ІСІН АВТОМАТТАНДЫРУДЫҢ ЛИНГВИСТИКАЛЫҚ НЕГІЗДЕРІ
- ҚАЗАҚ ТІЛІНІҢ КОМПЬЮТЕРЛІК ҚОРЫН (БАЗАСЫН) ҚҰРАСТЫРУ ЖАЙЛЫ
- ҚОРЫТЫНДЫ

АЛҒЫ СӨЗ

Табиғи тілдегі ақпараттардың басым көпшілігі сөйлеу тілі түрінде немесе жазба мәтін түрінде кездесетін болғандықтан, табиғи тілдік ақпаратты компьютер арқылы өңдеудің маңыздылығы артуда. Осы айтылғандарға сәйкес мәселелердің қолданбалы және теориялық лингвистикаға, қолданбалы филологиялық зерттеу аясына да қатысы бар. Оның қолданбалы филология саласына қатыстылығы – мәтінге түсініктеме беретін ғылымдар қатарына жататындығынан. Сонымен бірге қазіргі кездегі бұқаралық қатынас құралдары мен іс прозасының барынша даму сатысында оның маңыздылығы бұрынғыдан да арта түсуде.

Ақпараттың әрбір түрі өз мазмұны арқылы сипатталады да, бірақ олардың бәрі бірдей машиналық (компьютерлік) көрініске қатысты жалпы заңдылыққа бағынады. Сондықтан информатика ғылымының негізгі міндеті ақпараттың барлық түрін қазіргі техникалық құралдар (компьютерлер) арқылы дайындау, өзгерістер енгізу, ұзақ мерзімге сақтау, тұтынушы қауымға тарату мен оны қолданудың заңдылықтарын зерттеу.

Ғылым ретіндегі информатика саласының кең мағынадағы түсінігі бойынша, ол әлемнің (ғаламның) ақпараттық моделін зерттеу мен құрастыру мәселесімен айналысатын ғылым саласына айналып отыр. Информатика мәселелерінің негіздері тілдік және мәтіндік бірліктердің атқаратын рөлімен жанасып жатады. Бұл бірліктер ең кіші құрастырылымдардан бастап, олардың ең ірілерін қоса есептегенде барлық тілдік деңгейлерді қамтиды деуге болады. Тілдің әр деңгейіне және оның толық пішініне немесе ішкі жүйелеріне түсініктеме бере алатын әмбебап сипаттағы лингвистикалық теория әлі де жоқ және ондай теорияның болу-болмауының өзі де күмән туғызады. Сондықтан модельдеу мен зерттеудің нысаны ретінде тіл қызметінің әр аспектілері жеке қарастырылуы керек.

Енді компьютерлік лингвистика саласына тоқталатын болсақ, оны да тіл ғылымының осындай жеке бөліктерін зерттейтін ғылым саласы деуге болады. Компьютерлік лингвистика – информатиканың лингвистикалық негіздерін және тілдік байланыстардың барлық аспектілерін, ойлау мен ақиқаттықты модельдеу әрекеттерін компьютерлік программалар арқылы тікелей зерттеп танумен айналысатын ғылым саласы. Сол сияқты, компьютерлік тіл білімі **тілдік бірліктердің** сөзден кіші, сөзге тең, сөзден үлкен түрлерін, яғни сөз тіркестерін, сөйлемдерді (айтылымдарды), толық мәтін мәселелерін және тілдік әрекеттерді модельдеу проблемаларын зерттейді. Мысалы, соңғы айтылғанға қатысты, мәтін ішіндегі мән-мағынаны анықтау мен бір тілден екінші тілге аудару мәселелерін атауға болар еді. Лингвистикалық автоматтардың мүмкіндіктерінің күннен-күнге артуына байланысты компьютерлік лингвистиканың шешетін мәселелерінің де аясы кеңейіп, тереңдей түсуде. Оның шеңберінде жаңа әдістер мен теориялар тәжірибе жүзінде зерттеліп, сыннан өтуде.

Біз ұсынып отырған оқу құралында "қолданбалы лингвистика", "инженерлік лингвистика", "информатиканың лингвистикалық негіздері", "компьютерлік лингвистика" және т.б. терминдер қолданылады. Олардың әрбіреуіне жеке тоқталмай-ақ, бәрінің де "қолданбалы лингвистика" пәнінің аясына сыйысып кететінін, оның кең мағыналы екендігін айтуға болады. Қолданбалы лингвистика термині компьютерлік лингвистика мен математикалық лингвистиканы ғана біріктіріп қоймайды, сонымен бірге лингводидактика мен одан да басқа ғылымдар салаларын қамтиды. Жоғарыда аталған терминдер ішіндегі "информатиканың лингвистикалық негіздері" атты терминнің ең тар мағынаға ие екенін ескеру қажет. Оны, шын мағынасында, информатиканың компьютерлік және ақпараттық ғылымдар салаларымен тығыз қатыстылығынан туындайтын лингвистикалық мәселелердің жиынтығы деп те қарастыруға болады. "Компьютерлік лингвистика" терминін аталған екі ұғымның аралығындағы мағынаға ие деуге болады. Ол "информатиканың лингвистикалық негіздеріне" тән мағынадан кең түрде қолданылады. Дәлірек айтқанда, мысалы, ол лингводидактикаға көбірек қатысы бар оқыту-үйрету жүйесінің лингвистикалық мәселелерімен де айналысады. Бірақ "компьютерлік лингвистика" терминінің "қолданбалы лингвистика" терминіне қарағанда мағынасы тар.

Қазіргі кездегі компьютерлік лингвистика саласының кейбір қолданыс табу аясын осы оқу құралында қарастыруды мақсат еттік.

Компьютерлік лингвистика ғылымының құрамы мен құрылымы жайлы пікірлер әлі де болса тұрақтала қойған жоқ. Бірақ мұндай жағдай әр ғылым саласының да бастапқы кезеңіне тән екені белгілі. Сондықтан компьютерлік лингвистика туралы сөз болғанда оны қолданбалы лингвистиканың компьютерге, компьютерлік программалар жазуға, **жасанды интеллектіге** қатысты мәселелері деп түсінген жөн.

Информатиканың лингвистикалық негіздерін әртүрлі түсінікте құрастыруға болады. Мәселен, Ю.Н.Марчук "Основы компьютерной лингвистики"[1] атты оқу құралын жазуда өз алдына компьютерлік лингвистикаға байланысты "информатиканы" тілдік деректермен қамтамасыз етудің нәтижелерін жинақтау мен даму болашағын белгілеу мақсаттарын қойған.

Ал қолданбалы тіл білімі бойынша, ең алғаш 1996 жылы С.-Петербургте жарық көрген "Прикладное языкознание"[2] атты оқулықтың құрылымы мен мақсат-міндеттері жоғарыда аталған еңбектен басқаша. Онда тіл білімінің қолданбалы саласының негізгі мәселелері (әдістемеліктен жалпығылымдыққа дейін) дербес қарастырылған. Басқаша айтқанда, оқулыққа енген материалдар атқаратын қызметіне қарай топтастырылып, жеке-жеке тақырыптар бойынша көрініс тапқан. Мысалы, "Орфография", "Автоматизированные обучающие системы", "Психолингвистика", "Учебная лексикография", "Машинный перевод" және т.б.

Аталған оқулықтың оқыту процесіндегі теориялық және әдістемелік артықшылықтарын ескере отырып, біз ұсынып отырған "Компьютерлік лингвистикаға кіріспе" атты оқу құралында да сол құрылымдық және мазмұндық принциптер ұстанылды. Сонымен бірге, ол оқулықтың компьютерлік лингвистикаға қатысты тақырыптары автордың ұғынуынша қазақшаға аударылып, әл-Фараби атындағы Қазақ ұлттық университеті филология факультетінде "Қолданбалы

---

[1] Основы компьютерной лингвистики: Учебное пособие. М.: Изд-во МПУ "Народный учитель", 2000. 226 с.

[2] Прикладное языкознание: Учебник / Отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. 528 с.

лингвистика" мамандығы бойынша оқитын студенттерге (2001 жылдан бері) және Абай атындағы Қазақтың ұлттық педагогикалық университеті қазақ филологиясы факультетінің магистранттарына дәріс беруде пайдаланылғанын атауға болады. Оқылған теориялық материалдар қажетті жүйеге келтіріліп, ұсынылып отырған оқу құралына енді. Оның құрылымына келетін болсақ, ол алғы сөзден, кіріспеден, компьютерлік лингвистика саласының маңызды мәселелеріне қатысты 11 күрделі тақырыптан және қорытындыдан тұрады. Әрбір тақырыпқа қатысты пайдаланылған әдебиеттер тізімі әрбір тақырыпқа қатысты мақаланың соңында берілді.

Әрине, бұл тақырыптардың бәрі бірдей толық дәрежеде баяндалды деуден аулақпыз. Еңбектің негізгі мақсаты – жоғары оқу орындары студенттеріне компьютерлік лингвистика ғылымына кіріспе ретінде мәлімет беретін қазақ тіліндегі оқу құралын ұсыну. Ескертпе түрінде айта кететін жайт, оқу құралы мәтінінде кейбір орыс тіліндегі және орыс тілі арқылы енген салалық терминдер автордың өз ұйғаруымен қазақ тіліне аударылып берілді.

КІРІСПЕ

"Компьютерлік лингвистика" термині (computational linguistics), негізінен, тіл қызметін модельдеуге қатысты пайдаланылатын компьютерлік бағдарламалар мен тілдік деректерді өңдеу, сол үшін қажетті компьютерлік технологияны жинақтау және компьютердің араласуына қатысты басқа да тіл зерттеу жұмыстарымен байланысты қолданылады. Кең мағынада алғанда, тіл білімі аясына қатысты теориялық және практикалық мәселелердің шешімін табуға байланысты компьютердің араласу жағдайларының бәрін бірдей компьютерлік лингвистика саласына жатқызуға болады.

Адам баласының ақпараттық ортада өмір сүретіні белгілі. Ал бұл ортада қабылдаушының жауапты ақпарат бойынша дұрыс шешімге келуіне жағдай жасайтын ақпараттық технологияның түрлері көп. Олар: автоматтандырылған ақпараттық жүйе, бұқаралық коммуникация құралдарына жататын теле-радио байланыс жүйелері, компьютерлік (машиналық) аударма жүйесі, диалогтік жүйе және сұрақ-жауап, т.б. жүйелер. Мұндай жүйелердің саны мен сапасы күннен-күнге артуда. Ақпараттық қоғамдағы тілдік практика мен қазіргі кезеңдегі тілді зерттеу теориясының қолданылу аясы да өсіп келеді.

Тілді зерттеудегі компьютерлік лингвистика саласы қолданбалы тіл білімі атқаратын қызметтердің бір тармағы ғана деуге болады, себебі "қолданбалы тіл білімі" ұғымы көптармақты бола келе, күнделікті өмірде дәстүрлі және жаңаша мағынада кездесіп жүр. Мәселен, бұл ұғым дәстүрлі түсінік бойынша, тіл қызметі мәселелерінің практикалық әдістерін жасауды, яғни жазу мәселесін дамытуды, ауызекі сөйлеудің транскрипциялық жүйесін құруды, шетел сөздерінің транскрипция жүйесін жасауды, стенография тәсілдерін дамытуды, арнайы лингвистикалық сөздіктер шығаруды, тарихи ескерткіштердегі таңба-жазуларды айқындауды, ғылыми-техникалық терминологияны қалыптастыруды т.б. іс-шараларды қамтиды.

Қолданбалы тіл білімінің жоғарыда сөз болған дәстүрлі саласы қазақ тілінде ойдағыдай дамып, түркі тілдері ішінде көш бастап келеді деуге болады. Ал оның компьютерге байланысты жаңа саласы – "компьютерлік қазақ тіл білімі" әлде де кешеуілдеп дамуда.

Қолданбалы тіл білімінің "статистикалық лингвистика" (квантитативті лингвистика) тарауы бойынша ғылыми ізденістер ғалым Қалдыбай Бектаевтың атымен және оның ғылыми мектебімен тығыз байланысты. Қазақ тілінің әртүрлі стильдерінің мәтіндері бойынша, алғашында **электронды-есептеу машинасының** (ЭЕМ), кейіннен компьютерлердің көмегімен алынған жиілік сөздіктер (әліпбилі жиілік, жиілік, кері әліпбилі) тілдің құрылымдық жағының статистикасы жайында көптеген мәліметтер алуға мүмкіндік туғызды. Осылайша, қазақ тіліндегі бай статистикалық мәліметтер негізінде бірнеше кандидаттық және докторлық диссертациялардың қорғалғаны көпшілікке мәлім. Бүгінгі таңда А.Байтұрсынұлы

атындағы Тіл білімі институтының "компьютерлік лингвистика" бағыты бойынша қол жеткен ғылыми нәтижелері айтарлықтай деуге болады. Осылардың тек негізгілері деп мынадай ғылыми жұмыстарды атауға болады:

- көркем әдебиет, көсемсөз, ғылыми-көпшілік әдебиет пен оқулықтар мәтіндерінің, алғашында ЭЕМ-ның, кейіннен компьютердің жадына көптеп енгізіліп, олар бойынша арнайы компьютерлік бағдарламалар құрастырылуының негізінде әртүрлі жиілік сөздіктердің алынуы;
- М.Әуезовтің 20 томдық шығармалар жинағының компьютер жадына енгізілуі және соның негізінде жазушы тілінің жиілік сөздіктерінің жарық көруі;
- осы тәрізді жиілік сөздіктердің негізінде қазақ сөздерінің әр стильде және әр көлемді мәтіндерде қайталануының қосынды жиілігінің, сол мәтіндерді қамту статистикасының анықталуы;
- сөзтұлғалардың құрылымдық статистикасы мен мәтіндегі пайыздық салмақтарының әр сөз таптарына қатысты айқындалуы;
- қазақ тіліндегі негізгі сөз таптарының мәтін бойында кездесулерінің математикалық статистика мен ықтималдық теория заңдылықтарына бағыну-бағынбау жағдайларының зерттелуі;
- қазақ тілінің "ТІЛ – ҚАЗЫНА" атты автоматтандырылған компьютерлік картотекалық қорының іске қосылуы;
- қазақ тіліндегі мәтін мазмұнын оның тұрпатына қарай ашудың (формалдаудың) негізгі принциптерінің айқындалуы;
- Тіл білімі институтында жоспарлы тақырыптар бойынша түзіліп жатқан түсіндірме, екі тілдік, бір тілдік сөздіктер мәтіндерінің компьютер жадында "сөздіктер қорын" құрап, тұрақты сақталуы.

Осы аталған іс-шаралардың қатарын, компьютердің араласуымен орындалған жұмыстардың және компьютерлік тіл білімі саласына қатысты басқа да жұмыстарды атап өтуге болады.

Компьютерлік лингвистика бағыты түркі тілдері үшін тіл зерттеудің жаңа саласы деп есептелгенімен, үнді-еуропа тілдерінде дәстүрлі болып, күнделікті зерттеу мәселелеріне айналған компьютерлік лингвистиканың кейбір жайттарына тіл зерттеудегі жаңаша көзқарасқа қызығушы қауымға арнап, арнайы сөз етуді жөн көрдік.

Өткен ғасырдың 50-жылдарынан бастап "қолданбалы лингвистика" термині жаңа мағыналарға ие бола бастады. Ол күнделікті өмірімізге компьютерлік технологияның және автоматтанған басқару жүйесінің араласуы арқылы мәтінді автоматты түрде өңдеуге қатысты көптеген қолданбалы бағыттағы тілдік бірліктерді талдау (анализ) мен жинақтау (синтез) негізінде тілдің танылуы мен талдану заңдылықтарын белгілеу қажеттігінен туындайды. Орыс тілінде жарық көрген әдебиеттерде "қолданбалы лингвистика", "есептік лингвистика", "инженерлік лингвистика", "автоматты лингвистика", "математикалық лингвистика", "статистикалық лингвистика" терминдері бірінің орнына бірі қолданыла бастады. Әрине, бұл аталған терминдерді ортақтастыру дұрыс шешім деуге болмайды, себебі олардың қолданбалы тіл саласындағы мақсаты мен міндетінде және әдіс-тәсілдерінде өзіндік ерекшеліктері бар. Тілдің қолданбалы шеңберінің кеңею сипатын оның әртүрлі ғылымдар салаларымен түйісу мүмкіншілігінен деп түсіну керек (*Городецкий, 23*). Тіл қызметінің пәнаралық сабақтастық қасиеті тілдік теорияға тек қана оң әсерін тигізумен қатар, қазіргі тіл білімінің тұжырымдамалық бағытын (концептуальдық аппаратын) жаңартуға да мүмкіндік туғызды.

Қолданбалы лингвистика саласының кең мағыналылығын – тілдің құрылымы мен қызметіне қатысты ғылыми мағлұматтарды тілге қатыссыз пәндер саласында, адам баласының практикалық мүқтаждығынан және мұндай қызметтің теориялық негіздемесі ретінде қолданылуынан туындайтынын көруге болады.

Қолданбалы лингвистиканы функционалдық көзқарас тұрғысынан қарастырсақ, оны тіл қызметінің ең қолайлы әдіс-тәсілдерін ұсынатын ғылыми пән деп те түсінуге болады. Тілдің коммуникативті функциясының қолайлы жақтарын ұсынатын пәндер – аударма теориясы, машиналық (компьютерлік) аударма, ана тілі мен шет тілдерін оқытудағы ақпараттық-іздестіру жүйесінің теориясы мен практикасы, ақпараттық жасанды тіл жүйесін жасаудағы таңбалау теориясы.

Тілдің коммуникативті функциясының бір көрінісі ретіндегі әлеуметтік қызметі әлеуметтік тіл білімінде, тілдік жоспарлау мен тілдік саясатта, орфография мен орфоэпияда, әсерлік теория мен саяси лингвистикада көрініс табатыны белгілі. Сол сияқты тілдің когнитивті қызметінің ең қолайлы жақтары компьютерлік тіл білімінде, психолингвистика мен афазиологияда және квантитативтік лингвистика салаларында көрініс тауып жүр (*Баранов, 8*).

Компьютерлік лингвистика информатиканың тілдік негіздемесін қарастыруымен қатар "тіл" мен "ойлау" арасындағы байланысты және одан туындайтын тікелей болмысты компьютер арқылы модельдеуге қатысты мәселелермен шұғылданады. Компьютерлік лингвистиканы тілдік бірліктерге қатысты проблемалар да қызықтырады. Ондай бірліктер сөзден кіші, үлкен және оған тең де болуы мүмкін. Бұл жердегі сөзден үлкен деп отырғандарымыз – сөз тіркестері, сөйлемдер (айтылымдар) және толық мәтіндер. Аталып отырған салада тілді зерттеу кезінде қолданылатын амал-тәсілдерді модельдеу жолдары да қарастырылады. Мәселен, мәтін мазмұнын ашу жолдарын немесе аудармашының бір тілдегі мәтінді екінші тілге аудару жолдарын модельдеу мәселелері бұған мысал бола алады.

Сөйлеу мен ойлау процестерінің арасындағы байланыстардың аса күрделі философиялық мәселе екені мәлім. Біз олардың түрлерін көптеген айтылымдардағы, сөйлемдер мен мәтіндердегі тілдік мәліметтерді тұрпатына қарай (формалды) сипаттауда ұтымды пайдалануға болады демекпіз.

Адам баласының "ойлау" қабілетін модельдеу жағдаятында ғалымдардың назары көбіне лингвистика мен логика арасындағы байланысқа ауысып, тұжырымдау процесі ойлаудың анықтаушы және шешуші сипаты болып саналады. Қазіргі тіл білімінде тіл мәселесіне байланысты кейбір жайттарды логика саласының мамандары қарастырып, ал кейбір логикалық тұрпаттар (формалар) сырын ашу жағы тілшілерге жүктелуде. Бұл түсінікті де. Себебі, грамматикалық (семантикалық) және логикалық тұрпаттар әр көзқараспен қаралатын бір ғана нысанның түрлері болып келеді. Табиғи тіл адам баласын дағдылы ақиқат шындығынан уақытша ажыратып, тілдік тұлғалардың семантикалық мағынасымен анықталатын "ойлау" формаларымен әрекет жасайтын жаңа деңгейге ауысуына жағдай жасайды.

Бұл айтылғандардың барлығы да компьютерлік лингвистика үшін аса маңызды мәселелер. Себебі, адам баласының ойлау қабілетін тұрпатына қарай (формалды түрде), әртүрлі дәлдікпен қайталаудың мүмкіншілігі туады. Ал бұл жағдай математикалық модельдеу мен машиналық логиканың көмегімен іске асатын "ойлау" қабілеті бар "жасанды интеллект" жасап шығаруға мүмкіндік туғызады. Бұл жердегі "жасанды интеллект" атты жүйені – іс жүзінде "ойлау" қабілеті бар және "шығармашылыққа" қатысты мәселелердің шешімін таба алатын компьютерлік бағдарлама (программа) деп түсінген жөн.

"Жасанды интеллекті" іске қосу үшін ең алдымен адам баласының белгілі бір жағдайда шешім қабылдаудағы немесе жауапты мәселеге шешім іздеу кезіндегі ойлау процесінің сырын зерттеп тану қажет. Әсіресе, адамның ойлау процесіндегі көптеген сатылардың ішінен ең маңыздыларын бөліп алуды мақсат еткен жөн. Бұл жолда күрделі ойлау қызметінің құрылымы жағына мән бере отырып, шығармашылықтың құпиясын қарапайым сатылы бөліктерге бөлшектеп, компьютерлік бағдарламаның алгоритмін құрудың, болашақ "ойлау" қабілеті бар "жасанды интеллект" жасап шығарудың алғашқы кірпішін қалау жолдары зерттеледі.

Ми қызметін дәлме-дәл анықтау мүмкін бола бермейді, себебі адамға тән ақыл-ой – өте күрделі әлеуметтік-биологиялық функция, сондықтан да оның құпиясын ашу көптеген зерттеулерді қажет етеді.

"Жасанды интеллектіге" қойылатын негізгі талап ми қызметінің моделін "ойлау құралы" деп ұйғарудан туындайды. Тіл қызметіне қатысты модельдеудің негізгілері болып есептелетін – аударма мәселесі. Бұл тілдік құрылымның морфологиясынан синтаксисіне дейінгі барлық деңгейін қамтиды. Сондықтан да компьютерлік аударма тіл қызметінің аса күрделі функцияларының бірі болғандықтан, оны модельдеу де оңайға түспейді. Қазіргі кезде көптеген еуропа тілдерінен орыс тіліне (және керісінше) компьютерлік аударма жасау мәселесі біршама шешілді. Ал осы сияқты аудармаларды түркі тілдері бойынша, оның ішінде қазақ тіліне қатысты жүзеге асыру – егеменді еліміздің кезек күттірмейтін мәселелерінің бірі деуге болады. Қазақ тілінің компьютерлік саласы жаңа ғана дамып келеді және оның ойдағыдай дәрежеге жетуіне барлық мүмкіндік бар. Себебі, компьютерлік лингвистика саласы жеке ғылыми бағыт ретінде өткен ғасырдың 60-жылдарында ғана қалыптасты. Бұл сала бойынша жарық көрген ғылыми жұмыстар жеткілікті. Мәселен, АҚШ-та әр тоқсан сайын "Компьютерлік лингвистика" атты ғылыми журнал жарық көреді. Осымен байланысты компьютерлік лингвистика мәселесі бойынша айналысатын аймақтық құрылымдағы әлемдік "Қауымдастықтың" және әрбір екі жыл сайын осы сала бойынша өткізілетін "КОЛИНГ" атты халықаралық конференцияның жұмыстары да айтарлықтай жемісті деуге болады (*Баранов, 14*).

"Компьютерлік лингвистиканың" айналысатын мәселелерінің ең маңыздыларының бірі – табиғи тілдің мәтін морфологиясын автоматты түрде талдау. Егер, дәстүрлі түсінік бойынша, "сөз морфологиясы" деген ұғым сөздің тұлғалық құрылымын, яғни сөз түбірін және оған жалғанатын қосымшалардың (жұрнақ, жалғау) түрін анықтайтын болса, компьютерлік тіл біліміндегі "морфология" терминінен басқаша ұғым туындайды. Дәлірек айтсақ, ол сөздің мәтін бойындағы сыртқы тұрпатына қарай, тілдік құрылымның әрқилы деңгейлері бойынша тілдік мәліметтер алудың мүмкіндігін білдіреді. Алғашында "морфологиялық талдау" ұғымы айтылған мағынада **машиналық аударма** саласында пайда болды *(Василевский)*.

"Морфологиялық талдаудың" жаңаша түсінігі бойынша алынатын ақпарат біздің дәстүрлі ұғымымыздағы морфологияға қатысы жоқ көптеген амалдардан тұруы мүмкін. Осының салдарынан компьютерлік тіл біліміндегі "морфологиялық талдау" ұғымы – амалдар ұғымы (операционное понятие), яғни сөздің сыртқы пішіні бойынша тілдік ақпараттарды танып-білу үшін жүргізілетін іс-әрекет деп ұққан жөн сияқты. Басқаша айтсақ, егерде дәстүрлі тіл білімі ұғымы бойынша "морфологиялық талдау" кезінде "нені талдаймыз?" деп сұрақ қоятын болсақ, компьютерлік тіл білімінде "қалай талдаймыз?" деген сұрақ қойылуы керек, яғни

ол сөз тұлғасы бойынша қажетті тілдік ақпараттарды қай жолмен алуға болатындығымен айналысады.

Машиналық (компьютерлік) аударманың алғашқы тәжірибелерінде мұндай амалдардың саны көптеп кездесетін, ал бүгінгі таңда орыс және еуропа тілдері бойынша бұл мәселе біршама шешілді деуге болады (*Лингвистические...*). Ал қазақ тілі мен басқа да түркі тілдері бойынша мұндай ізденістер айтарлықтай қолға алынбай жүр.

Аталған түсініктегі "морфологиялық талдау" бірнеше бағытта қарастырылуда. Олардың біріншісі – сөзтұлғаны оның негізі мен оған жалғануы мүмкін болатын қосымшаларға ажырататын классикалық талдау негізінде модельдеу.

Екінші бағыт сөзтұлғалардың соңғы әріптерінің тіркесімдік заңдылықтарының жиілік сөздіктердегі статистикалық мәліметтеріне сүйенеді.

Үшінші бағыт – соңғы кездердегі ізденістердің нәтижесі. Бұл бағыт бойынша теңдеулердің ашық жүйелерінің пішіні ретінде морфологияның әмбебап математикалық моделі жасалады. Модельдің есептегіш мүмкіндігі негізінде сөзтұлғаларды нормалау және қажетті грамматикалық ақпарат алумен қатар сөзтұлғаларды жинақтау (синтездеу) да іске асады (*Марчук, 44*).

Сөзтұлғалардың әріптік құрамының өзгеруін анықтайтын сөздерді өзара топтауды (топтарға бөлуді) негіз етіп алып, автоматтандырылған морфологиялық талдаудың алгоритмі құрастырылады. Мұндай топтау "морфологиялық топтау" деп аталып жүр (*Белоногов, Новоселов*). Сөзтұлғалардың жазылуындағы әріптік өзгеріске ұшырау флективті және агглютинативті тілдерде біркелкі емес.

Мәселен, орыс тіліндегі: *"сижу – сидишь"* сөзтұлғаларының жазылуындағы сөз негізіндегі әріптер өзгеріске ұшыраса, сол сөздердің қазақша баламасы – *"отырмын – отырсың"* сөздерінің жазылуында қосымшадағы әріптер ғана өзгерген.

АМПАР атты машиналық аударма жүйесінде сөздердің синтаксистік қызметі мен септік және т.б. жалғаулардың негізінде сөздердің морфологиялық кластары екі түрге бөлініп қарастырылды:

1) негіздерінде өзгеріс болатын сөздер тобы,

2) сөздердің флективті тобы.

Соңғы топ белгілі сөздерге ғана тән қасиеттегі белгілер жүйесі арқылы немесе осындай қасиеттерді бойына сақтаған дерек сөздер арқылы сипатталады.

Енді морфологиялық талдаудың түрлеріне қысқаша тоқталайық. Олар төмендегідей:

- негіз сөздердің сөздігі арқылы морфологиялық талдау жүргізу;
- сөзтұлғалар сөздігі көмегімен морфологиялық талдауды жүзеге асыру;
- логикалық көбейту әдісін қолданып, морфологиялық талдау жасау;
- морфологиялық талдауды сөздіктердің көмегінсіз (сөздіксіз) арнайы кестелер арқылы жүзеге асыру.

Еуропа тілдерін зерттеуде көп тараған морфологиялық талдау түрі – негіз сөздердің сөздігі мен кейбір көмекші кестелер жүйесін пайдалану. Аталған сөздіктерде ішкі флексиясыз, жай және күрделі сөздердің негіздері мен олардың негіз тұлғалары (формалары) толық беріліп отырады. Сөздіктегі әрбір негіз сөзге морфологиялық кластардың екі түрін ажырататын шартты белгі (код) қойылады. Ал омонимдік негіздерге шартты белгілердің (кодтардың) тіркесімдік түрлері беріледі.

Морфологиялық талдаудың екінші түрі, яғни сөзтұлғалар сөздігі арқылы жүргізілетін талдау да зерттеушілер тәжірибесінде көп қолданыс тапқан, кең тараған талдау түрі болып саналады. Олай дейтініміз, морфологиялық талдаудың компью-

терлік алгоритмінде морфемаларға бөлшектеу мен оларды сөздік бойынан іздестіру әрекеттері негіз сөздер сөздігіне қарағанда жеңіл жүзеге асады. Бірақ талдаудың бұл түрінің өзіне тән осал жерлері де жоқ емес. Мысалы, егер іздестіретін сөз сөзтұлға сөздігінде кездеспеген жағдайда, біз оның грамматикалық ақпараты жайлы да ештеңе біле алмаймыз. Сондықтан сөзтұлға сөздігі көмегімен талдау жүргізу жүйесінде аффикстер мен түбір сөздер тізімдері берілуі қажет. Осымен бірге бірнеше сөзтұлғаларды бір **лексикалық бірлікке** сәйкестендіруге қажетті, оларға тән қасиеттерге ие белгілер (атрибуттары) толық берілгені талап етіледі. Бұл айтылғандардан туындайтын қорытынды – морфологиялық талдаудың негіз сөз сөздіктері арқылы жүргізілгені ұтымды.

Морфологиялық талдауда "логикалық көбейту" әдісі айрықша орын алады. Бұл әдісте "сөздік функциясы" деген ұғымға ерекше көңіл бөлу керек. Әрбір "сөздік функциясына" сөзтұлға функциясы мен оған тән ақпарат сәйкес қойылады. Осының негізінде әр сөзтұлғаның өзіне ғана тән ақпараттық деректер арқылы функцияның мәнін арнайы кестемен беруге мүмкіндік туады. Бірақ аталған функцияның мәнін басқаша, төменде көрсетілгендей тиімді амалдар арқылы беруге де болады:

1) сөзтұлға әріптер тізбегі ретінде морфемдік сегменттерге бөліктенеді;

2) сөзтұлға морфемдік сегменттер тізбегі ретінде басқа морфемдік элементтердің реттелмеген жиынымен ауыстырылады;

3) сөзтұлға морфемдік жиын ретінде белгілі-бір ақпарат сәйкес қойылады;

4) бұл ақпарат сөзтұлғаға қатысты деп саналатын қорытынды ақпарат түрінде қайта өзгертіліп беріледі (*Марчук,* 46).

Әрбір морфемаға құрылымында сондай морфемасы бар сөзтұлғалардың жиынтық ақпаратын да сәйкес қоюға болады. Логикалық түсінік бойынша, мұндай ақпараттар жиынтығы дизъюнкциямен, яғни сөзтұлғаның құрылымдық белгілерінің бір-біріне қарсы қойылуымен сәйкес келеді. Сөзтұлға жайлы ақпарат осы сөзтұлғаның табиғатына тән қиылысудан немесе логикалық конъюнкция деп аталатын морфемалық ақпараттардан тұратынын дәлелдеп шығуға болады.

Морфологиялық талдаудың логикалық көбейту әдісі, көбінде, флективті тілдерде қолданылады және негіз сөздер сөздігінің болуын қажет етеді.

Қарастырып отырған автоматтандырылған морфологиялық талдаудың соңғы түрі – сөздіксіз талдау немесе "тәуелсіз" талдау. Ол аффикстер кестелері көмегімен және арнайы түзілген грамматикалық мағынасыз сөздердің тізімі негізінде іске асады. Морфологиялық талдаудың бұл түрі зерттеу тәжірибесінде өте сирек қолданылады.

Енді "автоматтандырылған морфологиялық талдаудың қазіргі күйі қандай?" – деген сұрақтың жауабына қысқаша тоқталайық.

Қазіргі кезде кез келген мәтіннен компьютер арқылы сапалы ақпарат алу жүйесіне қойылатын талап заман сұранысына қарай өсті. Сондықтан компьютер көмегімен морфологиялық талдау жасау сызбасын (алгоритмін) жүзеге асыру мәселесінің негізі деп саналатын мынадай талаптар қойылады:

- кез келген тақырыптағы мәтіннің 98-99 пайызды қамтитын көптақырыпты (политематический) "аса күшті" сөздік морфологиялық талдау жүйесінің негізін құрауы тиіс;

- автоматты түрдегі талдау алгоритмі кез келген сөзтүрленістерін ескеретіндей мүмкіндікке ие болуы қажет. Осындай жағдай ескерілсе ғана сөзтұлғаны жанжақты тану мүмкіндігі артады және сан жағынан алғанда ондай мүмкіндіктің нәтижесі сөздікте қамтылған лексикалық бірліктен бірнеше есе артық болуы мүмкін;

- автоматтандырылған морфологиялық талдау жүйесінде "жаңа" сөздер, сөздікте қамтылған басқа сөздермен бірдей дәрежеде қарастырылуы қажет және оларды дұрыс танудың ықтималдығы 90-95 пайыздан кем болмауы тиіс;

- мәтінді компьютер арқылы өңдеу жылдамдығы тұтынушыны қанағаттандырарлықтай болуы қажет және оған енгізілетін мәтін көлеміне шек қойылмауы керек;

- компьютер арқылы жүргізілетін морфологиялық талдау жүйесінде, қажеттікке қарай, өзгерістер жүргізу (сөздікті толықтыру, өңдеу және т.б.) мүмкіндігі алдын ала ескерілуі тиіс, яғни бұл аталған жүйені "оқытуға" болатындай мол мүмкіндік тууы керек (*Марчук,* 48-49).

Бүгінгі таңда, орыс тіліне және басқа да еуропа тілдеріне қатысты морфологиялық талдау жүйелеріндегі кейбір сөздіктер мен мәтіндердің көлемдері жайлы мынандай мәліметтерді айтуға болар еді:

- негіз сөздерден тұратын көптақырыптық сөздіктің көлемі 100 мың лексикалық бірліктен кем емес;

- сөздіктің көлемі 30 млн сөзқолданыстан асатын көптақырыптық мәтін негізінде жасалғандықтан, оның кез келген тақырыптағы ғылыми-техникалық мәтінді қамту деңгейі өте жоғары;

- көлемі 3 млн сөзқолданыстан тұратын мәтіннің негізінде түзілген сөзтұлғалар сөздігі 46 мың лексикалық бірлікті құрайды.

Сөз болып отырған екі сөздікте де сөздердің құрылымы мен синтаксистік сипаты жөнінде толық түрдегі грамматикалық ақпарат берілген (*Зеленков*).

Автоматтандырылған морфологиялық талдау сөздіктердің айрықша түрлерін өмірге келтірді. Оларда тілдік бірліктердің тұлғасына, туындау ережелеріне, сөзтудырғыш, сөзтүрлендіргіш топтарына қарай реттеу жақтары қарастырылды. Осындай типтегі сөздікке тілдік бірліктің соңғы жағынан бастап әліпби тәртібіне келтірілген "кері әліпбилі сөздік" деп аталатын сөздікті жатқызуға болады. Морфологиялық талдаудың талабына сай түзілген бұл сөздікте сөзтудырғыштық, сөзтүрлендіргіштік типтегі сөздер және соңғы қосымшалары бірдей болып келетін күрделі сөздер өзара жіктелген ретте көрініс табады. Кері әліпбилі сөздік бойынша сөздердің морфологиялық құрылымына қатысты көптеген мәселелерді анықтауға болады. Мәселен, қосымшалары бірдей болып келетін сөздердің топталып берілуі олардың грамматикалық сипатын және қосымша мен қандай сөзтүрлендіргіш типке жататындығы арасындағы қатынастарға қатысты деректерді анықтауға мүмкіндік туады.

Грамматикалық форманттардың синонимдік, омонимдік қатарларын жеңіл ажыратуға және олардың тіркесімдік, сандық мөлшерінің сипаттарын анықтауға мүмкіндіктер туады. Кері әліпбилі сөздік бойынша, аналитикалық тілдерге қарағанда флективті және агглютинативті тілдер жүйелерінен морфологиялық форманттарға байланысты ақпараттарды көптеп алуға болады.

Осы уақытқа дейінгі қазақстандық "Статистика тобының" қазақ тілін ЭЕМ-ның және компьютердің көмегімен лингво-статистикалық зерттеу барысындағы қол жеткізген көптеген ғылыми нәтижелерін, яғни әр мақсатпен түзілген жиілік сөздіктер мен олар арқылы алынған құнды деректер мен тұжырымдарды әрі қарайғы іс-тәжірибемізде дұрыс пайдалана білуіміз керек. Ең алдымен, сол ғылыми жетістігімізді үнді-еуропа тілдеріндегі "компьютерлік лингвистика" саласы бойынша орындалып жатқан ғылыми-зерттеу жұмыстарының деңгейіне дейін көтеру қажет.

Компьютердің араласуымен ғана жүзеге асатын морфологиялық талдаулар мен басқа да зерттеу түрлерін мемлекеттік дәрежедегі қазақ тілі аясына қолдана отырып, ана тіліміздің компьютерлік лингвистика саласын дамытуды және соның нәтижесінде қол жеткізген жетістігімізді күнделікті өмір қажеттігіне пайдалану ісі қазіргі заман талабы болып отыр.

Баранов А.Н. Введение в прикладную лингвистику. Эдиториал УРСС. М., 2001.

Белоногов Г.Г., Новоселов А.П. Автоматизация процессов накопления, поиска и обобщения информации. М.: Наука, 1979.

Василевский А.Л., Марчук Ю.Н. Вычислительная лингвистика. Учебное пособие для студентов отделения прикладной лингвистики. М.: МГПИИЯ им. М.Тореза, 1970.

Городецкий Б.Ю. Актуальные проблемы прикладной лингвистики / Новое в зарубежной лингвистике. Вып. XII., 1983.

Зеленков Ю.Г. Морфологический анализ в системах автоматической обработки научно-технической информации: Канд. дисс. М.: ВИНИТИ, 1988. 145 с.

Лингвистические вопросы алгоритмической обработки сообщения. М.: Наука, 1983.

Марчук Ю.Н. Основы компьютерной лингвистики. Учебное пособие. М., 2000.

"АДАМ–КОМПЬЮТЕР–АДАМ" ЖҮЙЕСІНДЕГІ ТАБИҒИ ТІЛ ҚЫЗМЕТІНІҢ ЕРЕКШЕЛІКТЕРІ

Қоғамның даму сатысы күрделене түскен сайын, ондағы ақпарат құралы қызметін атқаратын "тілдің" дәрежесі де айтарлықтай көтерілді. Қоғамның даму сипаты ақпарат ағымының үздіксіздігі мен сапалылығына өз әсерін тигізбей қоймады. Әсіресе, қысқа және ұзақ мерзімді ақпарат түрлері айтарлықтай өзгеріске ұшырады.

Осының әсерінен, біріншіден, тілдік құрылымның негізгі қаңқасы ретіндегі грамматиканың, түбір сөз бен өнімді сөзжасам моделі жүйесінің тұрақтылығын, екіншіден, өндірістік қажеттілік пен қоғам өмірінің қажеттігіне сай сол қаңқаны толықтыратын лексикалық бірліктер мен олардың тіркесімдерінің сан жағынан ұлғайуымен бірге, олардағы тұрақсыздық қасиеттерді де байқауға болады (*Андреев*).

Адам өміріне электрондық техниканың араласуы бұлшық ет күшіне қатысты көптеген жұмыстарды жеңілдетумен қатар, ми қызметінің де іс-әрекеттеріне араласа бастағаны бәрімізге мәлім. Осыған байланысты адам миының қызметіне бәсеке тудыратын неше түрлі "ақылды" автоматтар мен "ойлау" қабілеті бар электронды машиналар жайлы жайттар соңғы кезде көптеген басылымдарда жарық көруде. Кейбір авторлар, тіпті осыдан 200 жыл бұрын тоқу станогының тоқымашыны ығыстырғаны сияқты, электронды машиналар да адамзатты ми қызметінің өрісінен ысырып тастап, оған тек тұтынушының ғана рөлін қалдыруы мүмкін деп те күмән туғыза бастады. Әрине, мұндай қисынсыз болжам адам баласының ми қызметінің мәнін дұрыс түсінбей, оны түгелімен бөлшектелетін (алгоритмделінетін) процеске жатқызып, компьютер арқылы оны қайта туындатуға болады-мыс деп ұйғарудан туындайды. Шындығында, адамның ойлау қабілеті негізінен дедуктивті-логикалық емес, индуктивті-ықтималдық болуынан, толығымен бөлшектелуге жатпайтын процесс деп ұйғару керек.

Бұл пайымдаулардан адамның ойлау қабілетін түгелімен формалдауға болмайтындығын және оны компьютерлік алгоритмдеу амалымен қайта туындатудың, яғни ойлау әрекетін қайтадан өзгеріссіз қалпына келтіру мүмкін еместігін ескеру қажет.

Соңғы кезде ақпараттар көлемінің өсуі жайлы жаңа көзқарас байқалады. Мәселен, ол – қысқа мерзімді ақпараттың көлемі арифметикалық прогрессия құрап өседі де, ал ұзақ мерзімді ақпарат – геометриялық прогрессия бойынша ұлғаяды деген көзқарастар. Сонымен қатар, ақпараттық процестерді компьютерлендіру (автоматтандыру) ғылым саласынан гөрі, өндіріс саласында көбірек және тезірек іске асады деп есептеу тенденциялары орын алып жүр.

Сол сияқты, табиғи тілге байланысты төмендегідей екі түрлі теориялық проблемаларды сөз етуге болады.

Олардың біріншісі – тілдің компьютер жадындағы қызметіне қатысты туындайтын сұрақ.

Табиғи тіл (ауызша не жазбаша) адам баласынан логикалық роботқа немесе компьютерге түгелдей ауысуы кезінде қандай сапалық өзгеріске ұшырауы мүмкін, әлде ондай "тіл" өзгеріссіз алғашқы қалпында қала ма?

Екіншісі – табиғи тілді зерттеу барысында формалдау процесі шекті ме? Басқаша айтқанда, "тілді" мазмұнға қатыссыз бөлек сипаттау мүмкін бе? Егер

олай қарастыруға болмаса, ондағы тілдік бірліктердің берілуі мен мағыналарының арақатынасын анықтау мүмкін бе? – деген сұрақтар туындайды.

Бірінші сұрақтың жауабын "тіл" ұғымын әртүрлі тұрғыда қарастыруға байланысты түсіндіруге болады. Мысалы, "тілді" тек семиотика тұрғысынан анықтасақ, онда тіл – символдарға қатысты кодтар жүйесі деп саналады. Бұл түсінік бойынша адам да, логикалық робот та кибернетикалық жүйе деп ұйғарылады. Осының негізінде "тілдің" адамнан роботқа немесе керісінше – роботтан адамға ауысу процесінде өзгеріс болмайды деп ұйғару керек.

Енді "тілді" мазмұнды лингвистикалық тұрғыдан қарастыратын болсақ, онда болмыста кездесетін "тілдің" екі түрі, яғни олардың жеке (дара) және ұлттық тіл түрлері болады деп ұғынған жөн. Тілдің кейбір қолданбалы мәселелерін шешуде семантикалық көзқарас ыңғайлы болғанымен, ол жеткіліксіз. Себебі табиғи тілдің екі жақтылық табиғаты оны таңбалар жүйесінің теориялық аясында толық баяндауға мүмкіншілік туғызбайды. Екі жақтылық табиғат, біріншіден, жекелік және ұлттық жүйе түрінде байқалса, екіншіден, коммуникативтік құрал ретінде көрініс табады.

Егер логикалық роботты (компьютерді) аса бір зор жазба кітапшасымен салыстыратын болсақ, оның ұқсастығымен қатар, айырмашылығын да аңғаруға болады. Ұқсастығы – екеуінде де табиғи тілдің жазба көрінісі болуында.

Компьютер мен адам арасындағы коммуникативтік жүйе арқылы мәтіндік ақпараттың адамнан компьютерге ауысу процесінде табиғи тіл оның көрінісі ретіндегі жазбамен алмасады да, ал керісінше, ақпарат компьютерден адам назарына ауысқанда, жазба мәтінінің компьютердегі көрінісі қайтадан таңбалардың психикалық кескін үйлесімінің бұрынғы қалпына келтіріледі, яғни ішкі тіл (внутренняя речь) қалпына ауысады.

Сонымен, кез келген мәтінмен компьютер арқылы жұмыс істеу – тілге психикалық әсер ету емес, ол мәтіннің формалды шартты белгілерінің оның өз таңбаларымен әрекет етуі деп түсіну қажет. Бұл әрекеттер, әрине, адамның ойлау қабілетіндегі процестерге ұқсай бермейді.

Табиғи тілдің құрылымын формалдаудың шегін анықтау қажеттігі тілдің берілу пішіні (план выражения) мен мазмұндық көрінісі (план содержания) арасындағы қатынастарын арнайы зерттеуді қажет етеді. Тілдік фактілердің формалдану байланыстарын зерттей келе, біз жазба тілдің берілу көрінісін баяндауда олардың ондай заңдылыққа бағынатынын байқаймыз. Ал тілдік элементтердің берілу көрінісі мен мазмұндық берілуінің арасындағы байланысты зерттеу барысында біз тікелей байқалмайтын нысандарға келіп тірелеміз. Бұл жағдайда "тілдік белгі" мен "белгіні білдіруші" өзара психикалық пішін үйлесімінде ғана байланысты болмай, олар мазмұнды тұрғыдағы үнемі өзгерісте болатын экстралингвистикалық ұғымдармен ықтималдық байланыстарға да ұшырайды. Тілдің мазмұндық болмысының толық формалдануына кедергі жасайтын, негізінен алғанда, тілдегі семантикалық құбылыстар.

Осы айтылғандардан мынадай қорытынды жасауға болады.

Адам баласына тән сөйлеу не жазу тілі еш уақытта логикалық робот – компьютерге толық берілуі мүмкін емес. Компьютерді тілдің берілу көрінісін анализдеу мен синтездеу және мүмкіндігінше оларды формалдау жағдаяттарында пайдалануға болады. Ал мазмұнға тиісті зерттеулер белгілі бір семантикалық шектен аспайтын көлемдегі формалдануға тиісті жайттарда ғана іске асуы мүмкін (*Андреев*).

Компьютер көмегімен шешілетін жоғарыда сөз болған тілдік көріністерді формалдауға көнетін есептерді квазилингвистика саласына жатқызады. Мұндай

есептер сан жағынан аз болғанымен, олардың қолданбалық шеңбері айтарлықтай мол болады. Квазилингвистикалық есептер адам өмірінің көптеген жақтарын қамтиды. Мысалы, олардың ойдағыдай шешілуі экономикаға байланысты жұмыстарды басқаруда, әскери қорғаныс істерінде және басқа да көптеген ғылыми зерттеу жұмыстарында маңызы зор деп саналады. Бүгінгі таңдағы осындай есептердің ең негізгісі – ақпараттық қор жинау. Бірақ бұл мәселені шешудің өзіндік қиындық жақтары бар. Ол, әсіресе, адам баласы мен компьютердің арақатынастарын (қызметін) дұрыс ажыратуға байланысты болып келеді. Мысалы, жиі қайталанатын және оңай формалданатын тілдік бірліктерге қатысты амалдарды: сұрыптау, топтастыру, санау, реттеу және т.б. жататын есептерді компьютерге жүктеген жөн болады. Ал стандарттауға жатпайтындар – семантикаға қатысты тілдік құбылыстар. Мысалы, мәтіннің мазмұнын ашуға қатысты – сөздердің көпмағыналығын айыру, омоним сөздердің мағынасын ажырату және сол сияқты кейбір формалдау процесіне көне бермейтін тілдік құбылыстар бар *(Андреев)*.

Енді мәтіннің туындауы (порождение текста) жағдаятын және оны әртүрлі деңгейде қабылдаудың мүмкіншілігі жайлы қысқаша мәлімет берейік.

Мәтін өзара қарым-қатынаста болатын тілдік жүйелердің "норма", "узус" деп аталатын ережелері мен шектеулеріне және тілдің өзіне байланыссыз пайда болатын сырт жағдайларына байланысты туындайды. Мұндағы "сырт жағдай" деп отырғанымыз, мәтінде баяндалатын оқиға мен шектеулерге (узусқа) қатысты контекстік жағдаяттар (*Пиотровский*).

Қабылдаушы адамның өзіне бағытталған хабардың мән-мағынасын анықтауы, оның тілдік жүйені меңгеру дәрежесіне және тілдің ережелері мен шектеулерін ("норма" мен "узус") айыра білуіне, әрі сол хабарламадағы сөз болатын жағдайлармен қаншалықты таныстығына да көп байланысты.

Қабылдаушы адам мәтіннің мән-мағынасын жан-жақты ажырата білуі, оның тілдік жүйені және оның "норма", "узус" деп аталатын қасиеттерін толық игеруімен қатар, осы айтылымда пайда болған жағдаяттың прагматикалық құндылығын ескеруіне де қатысты.

Енді хабарламаны лингвистикалық тұрғыдан қабылдау деңгейлерін сөз етер болсақ, онда мынадай сатыларды ажырата білуіміз керек:

1) грамматикалық деңгей (бұл деңгейде тек сөйлемнің грамматикалық схемасы ғана қабылданады);

2) сөздік деңгей (мұнда сөзтұлғалардың лексикалық мағынасы ғана қабылданады).

3) сөздік деңгейдегі жоғарғы саты – ол қабылдаудың фразеологиялық деңгейі. Бұл қабылдауда сөз тіркестерінің мағынасы ашылады, ал тіркес арасындағы грамматикалық байланыстар ескерілмейді.

4) келесі саты – жоғарыда аталған үш түрлі қабылдауды жалпылайтын саты – мәтін бойындағы **лексика-грамматикалық деңгейді** айқындау. Бұл сатыда да мәтін толық түрде лингвистикалық қабылдаудан өтпейді. Себебі қабылдаушы адам аталған деңгейлерде тек қана сөздер мен сөз тіркестерінің мағынасын ашып, мәтіннің грамматикалық құрылымын анықтағанымен, оның терең жатқан семантика-синтаксистік (әрі стильдік) байланыстарын аңғармай қалуы мүмкін. Мұндай байланыстарды анықтау үшін қабылдаушы адам мәтіндегі семантика-синтаксистік деңгейге көтерілуі қажет. Бұл деңгей, әрине, қабылдаушының тілді толық меңгеруін талап етеді.

Қысқаша айтқанда, болашақ автомат-роботтың тілдік білім дәрежесі әлеуметтік деңгейде болуы тиіс. Бірақ мұндай дәрежедегі электронды автоматты құрастыру

ісі, әзірше, көптеген күтпеген арақайшылықтарға ұшырауда. Осындай қиыншылықтардың негізгісі – адам баласының сөйлеу тіліндегі дискреттік (бөліктенулік) және үздіксіздік тәрізді қасиеттері болып табылады (*Пиотровский*).

Мәтінге қатысты инженерлік-лингвистикалық жұмыстар, яғни мәтінді толық қабылдау және оны "туындату" (порождение) тек "адам – компьютер – адам" (**А1–К–А2**) коммуникативтік жүйесінде ғана іске асады. Бұл жердегі адам, яғни "**А1**" – хабаршы, "**А2**" – хабарды қабылдаушы рөлінде, ал компьютер, яғни "**К**" – хабаршының хабарламасын қабылдаушы адамға жедел және қолайлы түрде жеткізетін аралық компонент.

"Адам – компьютер – адам" – хабардың берілу жүйесінің схемасы төмендегідей:

А1 ⟶ К ⟶ А2

Суреттегі **А1–К–А2** хабардың берілу жүйесін екіге: (**А1–К**) және (**К–А2**) бөліп қарастыруға да болады. Бірінші процесте (**А1–К**) мәтіндік хабарлама адамнан компьютерге беріледі, ал екіншісі – (**К–А2**), керісінше, компьютерден адам назарына қарай аударылады. Осы екі түрлі процестің әрбіреуін төмендегідей төрт блокқа бөліп қарастыруға болады (*Пиотровский*).

Бірінші процесс **А1–К**:

1.1) хабаршының (**А1**) назарында (ойлау процесінде) хабарлама мәтінінің "туындауы";

1.2) "туындалған" мәтін белгілерінің (кодтары) импульстер тізбегі ретінде компьютерге берілуі;

1.3) қабылданған импульстер жиынтығының компьютерлік семиозис негізінде мәтінді айқындау және оның мағынасын (мазмұнын) формалдау;

1.4) компьютер жадындағы айқындалған мәтінді мақсатқа сай өңдеу (өзгерту), мысалы, басқа тілге аудару, мәтіннен аннотация, реферат алу, индекстеу және нәтижесін табиғи тіл түрінде қабылдаушыға ұсыну.

Екінші процесс **К–А2** төмендегідей төрт топтамадан тұрады:

2.1) компьютер жадындағы мақсатқа сай өңделген ақпаратты тұтынушыға ұсыну (бұл топтама жоғарыдағы көрсетілген 1.4-ке ұқсас);

2.2) компьютерлік нәтижені қабылдаушы адамның (**А2**) қабылдауы;

2.3) компьютерлік хабарламаның семиозисі;

2.4) қабылдаушы адамның (**А2**) хабарламадан алған әсері.

Хабардың берілу жүйесі қызметінің ең маңыздылары деп, компьютерлік семиозисті (1.3), мәтінді айқындауды және қабылдаушының компьютер ұсынған нәтижеге байланысты түсінігін айтады. Осы аталған жағдайларда да семиозистің жетістігі табиғи тілдің компьютерлік тілге, яғни жасанды математикалық тілге ауысу нәтижесінің дәрежесіне тығыз байланысты.

Инженерлік тіл білімінің келесі бір маңызды проблемасы – лингвистикалық алгоритмдер мен олардың программаларын құру мәселесі. Мұны басқа сөзбен – "компьютерді оқыту" деп те айтуға болады. Табиғи тіл мен жасанды тілдің қатынастық жақтарын дұрыс түсінбейінше, бұл мәселені шешу оңайға түспейді.

Табиғи тіл де, жасанды математикалық (компьютерлік) тіл де семиотикалық жүйеге жататындықтан, олардың негізгі қасиеттеріне тоқталайық.

Тіл жүйесін тілдік бірліктер және олардың арасындағы қатынастар деп түсінетініміз мәлім. Ал табиғи тілдің бірліктері дегеніміз: дыбыс пен фонема,

әріп, морфа мен морфема, сөз бен сөзтұлға және олардың құрылымдық түрлері, сөз тіркесі, сөйлем және олардың қаңқалық схемасы. Енді жасанды компьютерлік тілдің бірліктері – әріп, цифрлар мен символдар және олардың шартты белгілері (кодтары), компьютерлік түбір сөз, сөздер және сөйлемдер. Бұл жердегі компьютерлік сөз не түбір сөз деп отырғанымыз, белгілі бір түрде шектелген әріп тіркестері, ал компьютерлік сөз тіркес және сөйлемдер дегендеріміз, алдын ала келісім бойынша шектелген компьютерлік сөздер тізбегі болып табылады.

Табиғи тілдің мынадай үш түрлі функцияларын атауға болады *(Баранов)*:

а) семантикалық және сигматикалық ақпаратты таратуға мүмкіндік туғызатын лингвистикалық символды қалыптастыратын тілдің номинативті функциясы;

ә) тілдік белгілерді өзара байланыста ұстайтын және мәтіндік ақпараттың синтактикалық қасиетін сипаттайтын тілдің предикативті функциясы;

б) хабардың берілу кезінде тілдің номинативті және предикативті функцияларын хабарлаушы (**А1**) не қабылдаушы (**А2**) жеке адамның (жеке ұйымның) қалауымен ұштастыра алатын прагматикалық функция.

Жасанды математика тілінде және оның барлық тармақтарында, яғни компьютерлік тілдің берілуінің өзінде-ақ предикативті функция біте қайнасып жатады.

Жасанды компьютерлік тілдің номинативті функциясы белгінің (знактың) мағынасымен (десигнатымен) шектеледі. Ал бұл жасанды математикалық тілде прагматикалық функция атымен жоқ деуге болады (*Пиотровский).*

Сонымен, хабардың берілу жүйесі (**А1–К–А2**) жұмысының нәтижелілігі, біріншіден, мазмұнды хабардың көлемдік мөлшеріне және екіншіден, көпшілік қабылдаушының дайындық дәрежесіне де көп байланысты демекпіз.

Қысқаша айтқанда, инженерлік тіл білімінің жалпы тіл теориясына қатысты атқаратын кейбір қызметтері осындай.

Андреев Н.Д. Язык в обществе, использующием кибернетические машины//Вопросы социальной лингвистики. Л.: Наука, 1969. С. 25-40.

Баранов А.Н. Введение в прикладную лингвистику. Эдиториал УРСС. М., 2001. 360 с.

Пиотровский Р.Г. Текст, машина, человек. Л.: Наука, 1975. 327 с.

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖҮЙЕСІН ҚҰРАСТЫРУҒА ҚАТЫСТЫ МӘСЕЛЕЛЕР ЖАЙЫНДА

Қазіргі түсініктегі **электронды-есептеу машина** (ЭЕМ) деген сөз тіркесі ең алғаш 1946 жылдан бастап (ЭНИАК, АҚШ) қолдана бастады және бұл есептеу техникасының қарқындап даму кезінің бастамасы еді. Міне, осы үрдіс қазіргі кезде де өз жалғасын табуда.

Бірінші есептеуіш машиналар көлемді арифметикалық есеп-қисаптарды орындауға арналып, ол машиналар программалар арқылы әрекет ететін ғажайып үлкен калькуляторлар тәріздес болды. Мысалы, осындай ЭНИАК машинасы баллистикалық кестелерді есептеу үшін жобаланған болатын. Бірақ өткен ғасырдың 50-жылдарының басында-ақ ЭЕМ-ды пайдалану мұндай шағын аямен ғана шектелмеу керектігі анық бола бастады, яғни есептеуіш машиналардың ақпарат түрінің саннан да басқаша түрлері бойынша әрекет ету (өңдеу) мүмкіндігі бары анықталды. Дәлірек айтсақ, практикалық тұрғыдан қарастырғанда, кез келген ақпарат (мәтін, сурет, дыбыстық сигнал және т.б.) ЭЕМ көмегімен өңделетін сандар арқылы кодталуына (шартты сандық белгімен ауыстырылуына) мүмкіндігі бары айқындалды. Бірақ бұл жайт практикалық әрекеттер үшін айтарлықтай маңызды емес еді. Оның себебі қазіргі кездегі есептеу техникасы мен оны пайдалану әдістерінің даму сипатына қарай ЭЕМ үшін қажетті деген программаны жазып, мысалы, қайсыбір мәтінді өңдеуге қатысты қойылған міндеттерді шешуге болады. Ал бұл мәтіннің қалайша кодталғаны (шартты белгіленуі) туралы мәліметті білуді тіпті де қажет етпейді.

Есептеу техникасының мүмкіндігін (яғни ЭЕМ арқылы сандық емес ақпаратты өңдеу мүмкіндігін) тез арада ұғынуға математикалық логиканың қарқынды түрде даму барысы себепші болды. Мысалы, ең алғашқы есептеу машиналарының пайда болуынан бұрынырақ кезде алгоритмдер теориясының (математикалық логиканың тарауы) Тьюринг машинасы, рекурсивтік функциялар, Марковтың нормалды алгоритмдері және т.б. нәтижелерге қол жеткен болатын. Сонымен, XX ғ. 50-жылдарынан бастап қазіргі кезге дейін **электронды есептеу машиналары** сандық емес ақпараттың да әртүрлі мәселелерін өңдеуде кең қолданыс тауып жүр.

Өткен ғасырдың 50-жылдарының өзінде-ақ ғалымдарды мынадай сұрақтар мазалайтын: машинаны адамша ойлатуға бола ма немесе, басқаша айтқанда, жасанды интеллекті құрастыру мүмкін бе?

Шынында да, тірі адам ағзасындай ойлау қабілеті бар жасанды құралды ойлап табу мәселесі адамның тарихи даму кезеңдерінде бірнеше рет қойылғаны мәлім. Мәселен, алхимиктердің гомункулусты алу әрекеттерінен бастап, ежелгі және ортағасырлық аңыздармен жалғаса отырып, Карел Чапек пен Айзек Азимовтың роботтарына дейінгі мәселелерді мысалға келтіруге болады. Бірақ тек ең алғашқы есептеу машиналарының пайда бола бастауынан-ақ бұл мәселе, абстрактылы философиялық сызба мен көркем шығармалар деңгейінде емес, белгілі дәрежеде болашақтың ісі бола тұра, ақиқаттықты ескеретін деңгейде нақтылық түсінік ала бастады. Бұл мәселе ең бірінші рет ағылшынның тамаша математигі А.Тьюрингтің 1950 жылы жарық көрген мақаласында нақты тұжырымдалып, жан-жақты сипатта қарастырылды. Мақала "Машина ойлай ала ма?" деп аталған болатын.

Жасанды интеллекті құрастырып шығуға бола ма? – деген сұраққа жауап алу, бір қарағанда, жеңіл болғанымен, ол аса күрделі мәселе еді. Оның күрделілігі

мынада: біз әлі де болса адам баласының ойлау жағдаятын дұрыс зерттей алмадық және ойлаудың жалпыға ортақ сипаттағы теориясы әлі де жоқ деуге болады. Оны айтпағанның өзінде, біз әлі "ойлау" деген не? – деп қойылатын сұраққа да жауап бере алмаймыз. Осы уақытқа дейін "ойлау" процесінің жасанды интеллект мәселелерін шешуге қатынасы бар философтарды, психологтарды, инженерлерді, тілшілерді және т.б. мамандарды бірдей дәрежеде қанағаттандыра алатын анықтама жоқ.

Мәселен, "ойлау" қабілеті бар жүйе, яғни жасанды интеллект жүйесі құрастырылды деп есептелік. Сонда оның "ойлай алатынын" немесе "ойлай алмайтынын" қалайша тексеруге болады? Интеллектуалдықты қалай өлшейміз? Осы мәселелерге байланысты А.Тьюрингтің жоғарыда аталынған тұңғыш мақаласында мынадай өлшем ұсынылып және ол ғылыми әдебиеттерде "Тьюринг тесті" деген атпен аталып кетті. Қысқаша түрде оны былайша тұжырымдауға болады: егер адам белгілі уақыт аралығында электронды машинамен қарым-қатынас жасап отырып, өзінің адаммен емес, машинамен "тілдесіп" отырғанын байқамай қалғанын сезінетін болса, тек сондай жағдайда ғана машинаның ойлай алатындығын мойындауға болады. Мұндай жағдайда адамның машинамен "тілдесуі" тікелей жүзеге аспауы мүмкін. Мысалы, ақпаратты жолдауда адам қайсыбір құралды (телетайпты) пайдаланады. Әрине, мұндай шектеу қою барынша табиғи екені белгілі. Себебі, ол интеллектуалдық өлшемі ішінен "сыртқы көрініс", "дауыс тембірі" және т.б. деп аталатын өлшемдерді (параметрлерді) жоққа шығаруы мүмкін.

Тьюринг тестісінің өзіндік субъективтілігіне қарамай, ол интеллектуалдық өлшем ретінде зор құндылыққа ие болды. Оның ең үлкен құндылығы – адамның ойлау әрекетімен салыстыру әдісін негіз етіліп алынуы. Әзірше ойлаудың жалпы теориясының болмауынан "бірдеңенің" ойлау қабілеті бар немесе жоқ деп мойындау үшін оны тек адаммен ғана салыстыра аламыз. Себебі, ол мұндай қасиетке әлем бойынша иегер жалғыз ғана тірі жан болып саналады.

Сонымен бірге аталған Тьюринг тесті барынша әмбебап қасиетке ие. Сұрақ және жауап әдісі практикалық тұрғыда адам баласының интеллектуалдық әрекетінің кез келген саласын қамтуға мүмкіндік туғызады. Тьюринг тестісіне қосымша шектеулер енгізе отырып, біз интеллектуалдық өлшемінің кейбіреуін ғана туындата аламыз, яғни ол жасанды интеллектің кейбір жеке міндеттерін ғана шеше алады деген сөз. Мысалы, біз машина мен адам арасындағы "сұхбаттың" белгілі пәндік сала аясынан ауытқымауын талап етуімізге болады. Сол сияқты арадағы қарым-қатынас тілінің лексикасы мен синтаксисіне де және т.б. шектеулер қойып, тілдесуді жүзеге асыруға болар еді.

Жасанды интеллектің қандай жеке міндеттерді шеше алатындығы жайлы Тьюринг тестісінің ешбір көмегінің болмауынан, зерттеу мәселелерінің аясы интуиция бойынша ғана анықталады. Мұндай міндеттер топтамасын әр зерттеуші өзінше анықтаса да, оның "тұрақты өзегінің" болатындығы белгілі. Олар – тәжірибе жүзінде, көпшілік мамандардың жасанды интеллект шешетін міндеттер санына қосатын мәселелері: есептеу әрекетін орындау, шешім қабылдау, бейнелерді тану.

Есептеу әрекетін жүзеге асыру. Бұл термин көптеген мәселелердің басын біріктіреді: жоғары алгебраның теоремаларын дәлелдеу мен орта мектептің физика оқулықтарындағы есептердің шешімін табудан бастап, күнделікті қажеттілікті өтейтін мәселелерге дейін. Мысалы, тіпті "қолымыздағы бар азық-түліктен қандай тағам жасауға болады" деген сұраққа жауап алуға дейін. Әрине, бұл жерде аталған есептердің нақты шешімін табу ғана емес, осы тәріздес есептердің жауабын іздеудің

жалпы механизмін табу, яғни есеп шығару барысының моделін табу жайлы сөз болып отыр.

Әзірше, негізінен, әртүрлі математикалық есептерді модельдеуге қатысты мәселелер зерттелуде. Себебі, мұндай математикалық есептер формалдау әдісіне тезірек көнеді. Сонымен бірге аталған мәселенің жалпылама шешімін табу үшін адамның пайымдау, қисынды қорытынды жасау мүмкіндіктерін зерттеп, түсініп алу қажет. Тек математикалық есептерді модельдеу үшін ғана формалды логика пәнінің әдістерін қолдануға болар еді. Әрине, егер мұндай модельдеуге математикалық логика әдістерінің "күші" жетсе. Себебі адам баласы қолданатын логикалық әдістер формалды логикадан тіпті де басқаша болып келеді. Бұл айтылғаннан адамдар формалды логиканы еш пайдаланбайды деген қорытынды шықпайды, бірақ олар нақты түрде қисынды өзгерістер жүргізу мақсатында тағы басқа да, әзірше белгісіз, "бірдене" қолданатыны анық байқалады. Мүмкін, тек формалды логика ғана үстемдік ететін әлем өте-мөте көңілсіз болатын шығар. Сонымен, егер математикалық логиканы барынша жақсы зерттелген ғылыми пән деп есептесек, онда "адам логикасы" тіпті де зерттелмеген сала деуге болады.

Шешім қабылдау мәселесі, негізінен, стратегиялық ойындар түрлерінде қарастырылады. Мысалы, шахмат, дойбы және т.б. ойындарды модельдеуге қатысты жұмыстар тек жұртшылықтың көңілін көтеру үшін ғана жүргізілмейді. Дұрысы, керісінше. Шахмат және басқа да ойындарды жүзеге асыратын коммерциялық бағдарламалардың шын мәніндегі мақсаты бөлек. Олай дейтініміз, "ойын ережелері" тек тақта бетіндегі шахмат фигураларын бір орыннан екінші орынға ауыстырып отыру үшін ғана емес екен. Оның түпкі мақсаты, мысалы, шикізат пен құрал-саймандарды өндіріс орындарына орналастыру, әскери сахнадағы іс-әрекеттер үшін полктер мен армияны жылжыту, финанстық жүйе бойынша қаржының орнын ауыстыру және т.б. сол сияқты мәселелердің шешімін табуға сай ұқсастықтар деуге болады.

Әрбір нақты жеке жағдайға байланысты ең жақсы нәтижеге әкелетін ережелерді таңдау негізінде **шешім қабылдаудың** ортақ сипаттағы әдіс-тәсілі анықталады. Осының нәтижесінде тек тиімді автоматтанған эксперттік жүйелерді құруға ғана қол жетіп қоймайды, сонымен бірге ойлау әрекетінің жалпы теориясына да белгілі дәрежеде үлес қосылады деуге болады.

Бейнелерді тану. Адамға өзін қоршап тұрған материалдық әлемді үнемі топтастырып отыру әрекеті тән екені мәлім. Адам әртүрлі нысандардың шексіз жиынын топ-топқа бөлшектеп қарастыруға дағдыланған. Мысалы, екі алманың өзара ұқсастығының аздығына қарамай, адам баласы, оны бір сөзбен – ***алма*** деп атайды. Ал ***алмұрттың алмаға*** өте ұқсас болуына қарамай, оны басқа сөзбен – ***алмұрт*** деп атайды.

Сол сияқты, демалыстан оралған жақсы таныс адамның өзгеріп – күнге күйіп, сақал-мұрты өсіп, қара көзілдірік киіп алғанына қарамай, біз оны басқа адаммен шатастырып алмай танимыз.

Әртүрлі адамдардың бір әріптің өзін ұқсастырмай әртүрлі жазғандарына қарамай, біз олардың бір әріп екендігін тани аламыз және бір ғана сөзді бірнеше адамның дыбыстау әрекетінің ерекшеліктеріне қарамай, біз оны да бір сөз деп ұғынамыз.

Екі түрлі нысанды теңдестіру үшін адам баласы қандай көзге шалына бермейтін белгілерге сүйенеді? Яғни, мысалы, екі заттың да алма екендігін анықтау үшін немесе кешегі көрген адам мен бүгінгі көріп тұрған адамның бір адам екендігін айқындау және т.б. үшін адам баласы қандай белгі-мүмкіндікті пайдаланады?

Ойлау барысында адам баласы нысандар мен олардың топтарының қайсыбір эталондық бейнесін құрастырады деп жорамалдауға болады. Тануға тиісті нысанды адам өз санасында орын алған бейнемен салыстыру нәтижесінде ол нысанды бейне топтарының біреуіне жатқызады. Бұл жердегі "бейне" сөзін қайсыбір психологиялық немесе нейрофизиологиялық термин деп ұғынудан гөрі, интуитивті түрде, адамның сол нысанға байланысты қайсыбір түсінігі немесе нысан топтары жайлы оған таныс бірыңғай сипаттамалар деп ұғынғанды жөн демекпіз. Әрине, нақты зерттеулерде "бейне" ұғымы дәлірек айқындалуы мүмкін.

"Бейне тану" мәселесінің шешімін табудың негізгісі – ол ең күрделі мынадай екі сұраққа жауап іздестіруді қажет етеді: эталондық бейне қалайша құрастырылады (яғни адам қалайша әлемді бөліктейді және топтайды) және эталонға қарағанда нысан бейнесінің қандай айырым белгілері мәнді, ал қандайлары мәнсіз?

Адам өзін қоршаған ақиқаттық жайлы ақпаратты әртүрлі арналар бойынша алуға мүмкіндігі бар – көру, есту, сезіну және т.б. Сондықтан адамдарда бейненің әртүрлі типтерінің қалыптасуы мүмкін болғандықтан, бейнені тану ісі бірнеше мәселелерге ажыратылады: көру бейнелерін тану, акустикалық бейнелерді тану және т.б.

Тьюринг тестісінің жасанды интеллект құрастыруға қажетті мәселелерді шешуге байланысты көмегінің болмайтындығы жайлы жоғарыда да айтқанбыз. Солай бола тұра, жасанды интеллект мәселелерінің Тьюринг тестінде аталған ең бір маңызды бағыты – адам мен машина (компьютер) арасындағы қарым-қатынасты (тілдесуді) табиғи тілде жүргізуді қамтамасыз ету. Бұл мәселенің маңыздылығы тек жасанды интеллекті құрастыру ісіне біршама жақындату мен ойлаудың жалпы теориясына белгілі үлес қосу ғана емес, сонымен бірге толық түрде шешілмесе де, оның көптеген практикалық қолданыста пайдаға асатындығында. Мысалы, ол **диалогтік ақпарат-іздестіру жүйесінде**, **басқару ісінің автоматты жүйесінде** және басқа да арнайы тілдерді үйренуді және арнаулы сөздіктерді қолдануды қажет ететін жүйелерді құрастыруда аса маңызды.

Сонымен, қолданбалы тіл білімінде маңызды деп саналатын жасанды интеллектің көптеген мәселелерінен адамның сөйлеу әрекетін модельдеу проблемасы, сөйлеуді талдау мен жинақтау мәселелері жеке қарастыруды қажет етеді.

Адам мен машинаның табиғи тілде қарым-қатынас жасау барысын былайша елестетуге де болар еді: машина мәтінді өз жадына енгізіп, қабылдап, өз тарапынан белгілі түрде тиісті әсер (реакция) етеді. Мұндай әсер құрастырылатын жүйенің міндеті мен мүмкіндіктеріне қарай әртүрлі болуы да мүмкін. Мәселен, ол синтезделінген машиналық мәтін түріндегі сұраққа жауап беру әрекетінің әсері болуы немесе жүйенің ішкі жағдайының өзгерісі (мәтіндегі ақпаратты жауапсыз қабылдау) мен қайсыбір жай ғана әрекет (роботтың өз иесінің бұйрығын орындауы) түрінің әсері болуы да мүмкін. Осындай мүмкін болатын әсерлердің ішінен біз тек сұрақ-жауап жүйесі мәселелерімен ғана шектелеміз.

Мәтінді қабылдау кезінде машинаға мәтін "ұғынықты әсер" етуі үшін ең алдымен машина оны "түсінуі" қажет болады. Осыған байланысты мынадай сауал туындайды: мәтінді "түсіну" дегеніміз не және машинаның мәтінді дұрыс түсінгенін қалайша анықтауға болады? Бұл аталған сұрақтар аса күрделі сұрақтарға жататыны белгілі. "Ойлау" (мышление) терминіне анықтама берудің қиындығы сияқты, "түсіну" терминіне де қатысты анықтама жоқ деуге болады. "Ойлау" мен "түсіну" терминдерінің бір-бірінен алшақ еместігі жайлы ғана жорамал жасауға болады.

Зерттеушілердің "түсіну" критерийін туындатуы Тьюринг тестісіне ұқсастық сипатта қарастырудың нәтижесі болуы да мүмкін. Мәселен, егер машинаның мәтінге қатысты әрекеті барабар және дұрыс сипатта болса ғана машина мәтінді түсіне алады. Бұл критерийдің субъективтілік сипаты (Тьюринг тесті тәріздес) айқын: машинаның мәтінге әсер ету әрекетінің дұрыс-бұрыстығын кім анықтайды? Сонымен бірге аталған критерийде "түсіну" ұғымында "дұрыс түсіну" жайлы да сөз болып отыр. Яғни бұған "бұрыс түсіну" жағы да қатысады деген сөз. Қалай болғанда да жасанды интеллектің практикалық жүйесін құру жағдайында бұл критерийдің қажеттігін терістеуге болмайтындығы белгілі.

Мәтінді дұрыс түсінуге болатындығы жайлы аталған критерий де (Тьюринг тестісі сияқты) көмекші құрал бола алмайды. Бұл мәселенің шешімін табуға байланысты әр зерттеушінің өз ұстанымы болатындығы белгілі. Сонда да бұл бағытқа жалпылама түрде былайша сипаттама беруге болады: жасанды интеллект жүйесі мәтінді қайсыбір ішкі көрініс құрылымына қайта өзгертуі қажет, яғни мәтін мағынасы құрылымының көрінісіне келтіруге байланысты өзгертулер жүргізіледі. Осы құрылымды негіз етіп ала отырып, аталған жүйе өзінің "интеллектуалдық қызметін" атқара алады. Мәселен, қойылған сұраққа қатысты қажетті деген жауапты ұйымдастыру, қайсыбір әрекетті жүзеге асыру және т.б.

Ішкі көрініс құрылымының көпшілік мақұлдаған ешбір де ережелері мен нұсқаулары жоқ. Бар болғаны тек бір ғана барынша құңгірт сипаттағы талап – ішкі көрініс құрылымының мәтін мән-мағынасын дұрыс ашу мен әртүрлі "интеллектуалды" әрекеттерді жүзеге асыру мүмкіндіктеріне ие болу. Сонымен бірге әртүрлі мән-мағынаға әртүрлі құрылымның сәйкес келуі.

Кейбір зерттеушілер жасанды интеллекті жүйесінің мәтінді түсіну барысы "адамның түсіну" деңгейіндей болуын талап етеді. Ондай ғалымдардың түсінігінше, басқаша жағдайда табысқа жету мүмкін емес. Әрине, бұл талап та өзінше қисынды деуге болады. Бірақ бүгінгі таңда мәтінді дұрыс түсіне алатындығын мақтаныш ете алатан тірі жан иесінің бірі және бірегейі – ол тек адам. Сондықтан адам баласында ғана орын алатын мұндай жағдаят (процесс) осы сала бойынша оң нәтижеге әкелетін жалғыз-ақ мүмкіндік деуге болады. Бірақ жасанды интеллекті жүйесіне адам мүмкіндігіне сай келетіндей талап қою ақиқат шындыққа жараса бермейді. Әрбір зерттеуші адам өз түсіну мүмкіндігін сенімді түрде негіздей алмайды. "Адамның басы" зерттеушілерге қайсыбір "қара жәшік" тәріздес. Ал адам басындағы түсінік алу кезінде болатын жағдаяттарды барынша дәл айқындауға нейролингвистика ғылымының жетістіктері әлі де болса жеткіліксіз. Сондықтан мамандар өз зерттеулерінде аталған мәселенің "адам басында қалайша шешім табылатынын" жүзеге асыруға ниеттенгендерімен, олар сөйлеу қызметін функционалдық тұрғыда модельдеу қажеттігі жайлы аса сақтықпен айтуда.

Сонымен, жасанды интеллект жүйесі мәтінді "түсіну" үшін ол ең алдымен мәтінді ішкі көрініс құрылымына ауыстыруы қажет. Сонда ғана мәтіннің мән-мағынасы қажетті түрде бейнеленеді. Мұндай ауыстыру жұмысын жүзеге асыру үшін жасанды интеллекті жүйесі қайсыбір ақпаратқа сүйенуі керек болады. Әдетте жасанды интеллектімен айналысатын мамандар бұл ақпаратты ***білім*** деп атап жүр. Көпшіліктің мақұлдауы бойынша *білім* екіге бөлініп қарастырылады – *тіл туралы білім* және *әлем туралы білім*. Білімнің соңғы түрін *энциклопедиялық білім* деп те атайды.

Аталған *білім* түрлері аралығына дәл шекаралық сызық жүргізу барынша күрделі, тіпті мүмкін де емес деуге болады. Әрбір зерттеуші бұл мәселені өзінше шешуге тырысуда. Көбінде тіл туралы *білімге* морфологиялық, синтаксистік,

семантикалық ақпараттарды жатқызады. Бұл аталғандар мәтіннің құрылымы мен сөздер аралығындағы негізгі байланыстар арқылы мәтін құрылымын анықтауға мүмкіндік туғызатын мәліметтер. Ал мәтінді түсінуге қажетті барлық басқа ақпараттар *әлем туралы білім* түріне жатады.

Табиғи тілдегі мәтінді түсінуге арналған алғашқы эксперименттік жүйелерді құрастырушылар тек *тіл туралы білімдермен* ғана қанағаттануға талаптанған болатын, бірақ көп ұзамай-ақ мұның жеткіліксіз екендігі, яғни *қоршаған орта жайлы білімнің* де қажеттігі айқындалды.

Әртүрлі адамдардың *әлем туралы білімі* айтарлықтай дәрежеде айырым табуы мүмкін. Сол сияқты мұндай жағдай әртүрлі жасанды интеллект жүйелеріне де қатысы бар. Ал бұл айтылғандар мәтінді түсінуге өз әсерін тигізбей қоймайды. Сондықтан да әр адамның бір ғана мәтінді әртүрлі (өзінше) түсінуі орын алады. Бұл айтылғандарға мәтін үзінділерінен көптеп мысалдар келтіруге де болады. Міне, мұның өзі жоғарыда сөз болған критерийдің субъективтілігінің тағы бір дәлелі деуге болады.

Тіпті ең қарапайым деген мәтінді дұрыс түсіну үшін айтарлықтай көп көлемдегі *әлем жайлы білімді* пайдалануымыз керек болады. Егерде *тіл туралы білімді* азды-көпті дәрежеде толық сипаттауға мүмкіндік туатынына қарамай, *әлем туралы білім* қорының шексіздік сипаты оны толық қамтуға мүмкіндік туғызбайды. Солай бола тұра, адам баласының әлемді тану мүмкіндігі де күннен-күнге артып келетінін ескеру қажет. Осыған байланысты түсіну жүйесін құруда *әлем туралы білімге* қатысты кейбір принциптік мәселелер туындайды. Мәселен, барлық білім қорын қамту мүмкін болмағандықтан, жұмыс барысында олардың ең қажеттілерін алдын ала болжау да аса күрделі. Құрастырылып жатқан жүйелердің негізгі бөлігі аса шағын пәндік салаға бағышталған, сондықтан ондай жүйелер тек шектеулі тақырыпты (мазмұнды) мәтінді ғана түсіне алады. Бірақ осындай шектеу жасаудың өзінде де алынған нәтижелер шындықтан әлі де болса алшақ.

Аталған мәселеге қатысты келесі бір зерттеу бағыты – оқыту жүйелерін (обучающие системы) құруға байланысты. Бұл бағыттың болашағының зор екені белгілі. Себебі, оның түсінік алу жүйесі арқылы мәтін ішінен жаңа ақпаратты тауып, бұрыннан жинақталып жатқан *әлем туралы білім қорына* қоса алатындай мүмкіндігі бар. Егер мәтінді түсінуге қажетті ақпарат жеткіліксіз болып жатса, оның күңгірт (түсініксіз) жерлерін анықтау үшін ондай жүйе тұтынушы адамға қосымша сұрақтар туындатып, білім қорын жинақтауға мүмкіндігі бар. Бірдей көлемдегі білім қоры бар жүйелер алғашқы кезеңде әртүрлі нәтижелерге ие болуы мүмкін.

Бірақ әлем туралы білім қорын жинақтау жолдарына байланыссыз (тәуелсіз), мәтінді түсіну кезіндегі негізгі міндет – мәтіннің ішкі көрінісінің құрылымын, яғни мәтіннің мән-мағынасының құрылымын құрастыру. Әдетте мұндай құрылым мәтінді автоматты түрде өңдеу кезіндегі әрекеттерге ұқсас болып келеді. Мәселен, машиналық аударма, автоматты рефераттау және т.б. орын алатын морфологиялық, синтаксистік және семантикалық талдау әрекеттеріне аса ұқсас болып келеді.

Құрастырылған құрылымды жүйенің қарсы әрекетін туындату үшін де қолдануға болады. Оған жататындар: жауапты тұжырымдау, әрекет түрін таңдау және т.б. Әрине, аталған кезеңге қатысты зерттеулерді жүзеге асыру тек тілшілердің үлесіне ғана тиісті емес, сонымен бірге басқа да көптеген пәндер өкілдерінің тікелей араласуын қажет етеді. Мысалы, психологтардың – әрекет үлгісін анықтауда, логика мамандарының – қисынды қорытынды жасауда және т.б. жасанды интеллект

саласындағы барлық зерттеу бағыттары практикалық тұрғыда әртүрлі мамандықтардың мүдделеріне жақын келеді, сондықтан олар әртүрлі пәндік мамандарды қатыстыра отырып, кешенді түрде жүзеге асуы қажет.

Жоғарыда сөз болған мәселелердің бәрін бірдей қанағаттандыратын ортақ шешім әлі де болса табылмай келеді. Әрине, жеке бір мәселеге қатысты аса тар шеңбердегі мәселелердің шешімдерін табуға арналған талпыныстар көптеп кездесетіні белгілі. Мәселен, егер адам мен компьютер аралығындағы тілдесу табиғи тілде жүргізілетін болса, онда ол шектеулі пәндік сала бойынша, ал егер нақты шешім қабылдау үшін модельдеу әрекеті қажет болса, онда ол нақты мәселенің шекаралығында (шахмат ойынындағы жүрісті таңдау) жүзеге асады және т.б.

Сонымен, бүгінгі таңдағы ойлау әрекетінің жалпы теориясы жаңадан ғана дамып келеді деп қорытынды жасауға болады.

Моделирование языковой деятельности в интеллектуальных системах / Под ред. А.Е. Кибрика и А.С. Нариньяни. М., 1987.

Панков И.П. Искусственный интеллект//Прикладное языкознание: Учебник / Отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. С. 91-99.

Попов Э.В. Общение с ЭВМ на естественном языке. М., 1982.

Фитиалов С.Я. Понимание естественного языка в системах искусственного интеллекта//Лингвистические проблемы функционального моделирования речевой деятельности. Вып. V. Л., 1982. С. 3-9

Хант Э. Искусственный интеллект. М., 1978.

ОҚУ ЛЕКСИКОГРАФИЯСЫНЫҢ ҒЫЛЫМИ ПӘН РЕТІНДЕГІ ЕРЕКШЕЛІГІ

Егемен еліміздің орта және жоғары оқу орындарында салалық пәндерді мемлекеттік тілде оқыту аса күрделі және кезек күттірмейтін көптеген мәселелерді шешуді талап етіп отыр. Осыған байланысты басты мәселенің бірі – сөздіктер, оқулықтар мен оқу құралдарын өмірге келтіруде ғылыми педагогикаға негізделген қазақ тілінің жалпы лексикалық қорын және салалық пәндердің терминологиясын құрастыру.

Қазақ тіліндегі әр салалық пәндерге қатысты терминдерді қалыптастыруда ғалым-лексикографтарды орыс және басқа да туыстас түркі халықтарының озық тәжірибелеріне сүйенетін күрделі жұмыстар күтіп тұр деуге болады. Атап айтар болсақ, халық ауыз әдебиетінің бай мұрасы мен қазақ халқының тұрмыс-тіршілігіне, салтына тән көптеген ежелгі атаулар мен терминдерді оқырманның жас ерекшелігіне, білім деңгейіне қарай сөздіктер мен оқу құралдарында орынды пайдалана білу ісі жүзеге асар еді.

Қай ғылым саласын алмасақ та, оның өзіне тән метатілі болатыны мәлім. Метатіл ұғымына тек терминдер ғана емес, жалпы ғылыми лексика да енеді. Ал мұндай метатілдің қалыптасуына сөздіктердің де әсері болатындықтан, көпшілік ғалымдар тіл білімінің лексикография саласын теориялық және практикалық негіздемесі бар ғылыми пән ретінде қарастыруды жөн санайды.

Ахмет Байтұрсынұлы мен Құдайберген Жұбанов бастап дамытқан қазақ тіліне қатысты терминдер белгілі дәрежеде қалыптасқанымен, басқа ғылыми пәндердің қазақша терминдік жүйесі әлі де жинақтауды, зерделеуді қажет ететіні белгілі. Сондықтан да оқулықтарды, сөздіктерді пайдалана отырып, олардың түсініктемесін, түсіндірілуін айқындау қажет болып отыр.

Қазіргі кезеңде қазақ тілінің сөздік қоры жаңарып, тіл білімінің арнасы кеңеюіне байланысты оқулықтар мен сөздіктер де уақытқа сай өзгеріске ұшырауда. Жарық көріп жатқан сөздіктерде соңғы жылдары пайда болып жатқан қазақ және орыс тілдеріндегі жаңа сөздер мен байырғы сөздердің түсіндірмелері мен қолданыстары енгізілуде. Сол сияқты ондай сөздіктерде қазақ әдеби тілінің нормаларына сәйкес және күнделікті қарым-қатынаста, оқыту жүйесінде жиі қолданылатын сөздер де орын алуда. Айтылуы қиындық тудыратын қазақ немесе шет тілінің сөздері мен сөз тіркестерінің транскрипциясы беріліп, реестр сөздерінің стильдік ерекшеліктері қалыптасқан ескертпелер мен шартты белгілер арқылы көрсетіп беру жұмысы да атқарылуда. Әртүрлі мақсатты ұстаным ететін сөздіктерде сөздің қолдану сипатын жан-жақты қамту үшін кеңінен таралған қазақтың мақал-мәтелдері, сөздердің негізгі және ауыспалы мағыналарын айқындайтын сөз тіркестері келтіріліп, омонимдер мен көпмағыналы сөздер қатарларына ерекше көңіл бөлу ойластырылуда.

Әрине, қазақ тілінің лексикография саласының дамуына тіл білімінің төрінен орын алған мынадай іргелі еңбектер негіз болғаны мәлім: "Қазақ тілінің түсіндірме сөздігі" (1-10 том, А., 1974–86), І.Кеңесбаевтың "Қазақ тілінің фразеологиялық сөздігі" (А., 1977), Н.Т.Сауранбаевтың жетекшілігімен жарыққа шыққан "Казахско-русский словарь" (М., 1954), Р.Сыздықтың жетекшілігімен жарық көрген "Жаңа атаулар" (А., 1999), А.Байтұрсынұлы атындағы Тіл білімі институты лексикографтарының құрастыруымен және жалпы редакциясын Т.Жанұзақовтың басқаруымен шыққан "Қазақ тілінің сөздігі" (А., 1999), "Қазақша-орысша сөздік" (А., 2001), "Орысша-қазақша сөздік" (А., 2005) және т.б.

Қазақстан Республикасы Үкіметінің жанындағы Мемлекеттік терминологиялық комиссиясының әр жылдары ресми бекіткен түрлі сала терминдері жинақталып, дүркін-дүркін жеке сөздік түрінде шығарып тұру да дәстүрге айналды. Осындай лексикографиялық жұмыстардың нәтижесі ретінде 1976, 1979, 1990 жылдары тілші ғалым А.Әбдірахмановтың құрастыруымен шыққан "Атаулар мен терминдер" сөздігі, 1996 жылы жарық көрген Б.Қалиев пен Ш.Құрманбайұлының "Орысша-қазақша және қазақша-орысша терминдер сөздігі", Тілдерді дамыту департаментінің 1998 жылы Т.Алпысбаев пен А.Әбдірахмановтың және 1999 жылы жарық көрген құжаттар жинағының басын қосып, 2002 жылы басылып шыққан "Терминологиялық жинақ", сондай-ақ салалық терминологиялық орысша-қазақша, қазақша-орысша сөздіктердің 31 томы "Рауан" баспасынан жарық көрді (2000 ж.). Осы аталғандардың қатары-на 1998 жылы "Ғылым" баспасынан шыққан Б.А.Жұбановтың 25677 сөзді қамтитын көптілді (ағылшын, неміс, орыс және қазақ тілдері бойынша) "Химиялық сөздікті", А.Салқынбай мен Е.Абақанның авторлығымен "Сөздік-словарь" баспасынан 1998 жылы шыққан "Лингвистикалық түсіндірме сөздікті", Т.Момынов пен А.Рақышевтың 1999 жылы "Атамұра" баспасынан жарық көрген орысша-қазақша "Медициналық сөздікті" жатқызуға болады. Сол сияқты М.Қазбеков пен Р.Шойбековтың "Русско-казахский словарь" атты сөздіктің (А., 1996), 2002 жылы "Аруна" баспасы шығарған 30000 сөзді қамтитын мектеп оқушылары мен жоғары оқу орындары студенттеріне арналған "Сөздіктің" (10000-нан аса сөз, құрастырғандар: Б.Исмағұлова, Э.Ережепова, Г.Әбдіжапарова) және 2006 жылы Астанада жарық көрген З.Ө.Жантекеева мен О.Нақысбековтың "Қазақша-орысша және орысша-қазақша оқушыларға, студенттерге арналған сөздіктің" (14000 сөз) де қазақ және орыс тілдерін оқып-үйрену кезінде аса құнды лексикографиялық еңбектер екені белгілі.

Осы бағыттағы аталған және басқа да аталмай қалған көптеген басылымдардың қазақ тілін оқыту процесінде, әсіресе жоғары білімді, білікті мамандар дайындауда игі көмегін тигізері сөзсіз. Қазіргі жағдайда оқу процесін қамтамасыз ету үшін арнайы маман-педагогтар мен ғалым-лексикографтардың араласуымен қазақ тілінің оқулықтарын, оқу құралдарын, көптілді аудармалы, түсіндірмелі және энциклопедиялық сөздіктерін сапалы құрастыру заман талабы деуге болады. Өкінішке қарай, мектеп оқушылары мен студенттерге арналған қолымызда бар бірқатар сөздіктерде, оқулықтар мен оқу құралдарында әлі де болса теориялық тұрғыда кеткен кейбір олқылықтар та жоқ емес.

Сондықтан оқырман назарына ұсынып отырған еңбегімізде осындай кемшіліктерге болашақта жол бермеу тұрғысынан оқу лексикографиясына қатысты кейбір теориялық және практикалық мәселелерді сөз етуді жөн көрдік.

Оқу лексикографиясы жалпы лексикографияның жас саласы болып саналатыны белгілі. Солай бола тұра, оның да өзіне тән жақсы зерттелген тарихы мен дәстүрі бар. Ресейде ең алғашқы оқу лексикографиясының орталығы тек 1966 жылы ғана құрылып, Мәскеудегі А.С.Пушкин атындағы Орыс тілі институтында сектор ретінде қызмет етеді. Қазіргі кезде аталған сектор орыс тілінің біртілдік және екітілдік (орысша-шеттілдік) сөздіктерін құрастырудың теориялық және практикалық жұмыстарын кең көлемде атқаруда.

Оқу лексикографиясының ғылыми пән ретіндегі өзіне тән ерекшелігі, оның жалпы лексикография, лексикология, семасиология психо- және әлеуметтік лингвистика және т.б. салалардың тілді үйрету әдістемесіне қатысты мәселелерімен тығыз байланысты болуында. Сондықтан оқу лексикографиясы дәстүрлі жалпы лексикографияның мәселелерімен қатар, мынандай көптеген әдістемелік факторларды

ескеруді қажет етеді: оқыту кезеңдеріне қатысты сөйлеу әрекетінің түрі, тілді үйренушілер үшін тілдің ана тілі (немесе аралық-тіл) екендігі және қажетті тілдік ортаның болу-болмауы, жас ерекшелігі мен жалпы мәдени деңгейге қатысты болуы және т.б. мәселелер.

Осы аталған лингвистикалық және әдістемелік факторлар тек бірдей (және кешенді) дәрежеде ескерілсе ғана оқу лексикографиясын пайдалануға байланысты, ана тілін не шет тілдерін үйретудегі оқу сөздіктеріне қатысты дұрыс әдістемелік нұсқау беруге болады. Лексикографияның жалпы теориясы принциптерін орынды пайдалану негізінде тілді үйрету жолындағы ескірген әдістерге жол бергізбей, ішкітілдік және тіларалық баламалардың шегін анықтауға, лексика-семантикалық тіркесім және т.б. жайттарды айқындауға мүмкіндік туындайды.

Әдістемелік факторларды ескеру негізінде мына жайттар анықталады (*Марусенко,* 1996, 309):

1) сөздіктің адресаты (мектеп оқушысы, студент және т.б.);

2) сөздіктің типі (біртілдік немесе аударма);

3) сөзтізбе көлемі (минимум сөздік, орта немесе толық);

4) сөздік мақаланың құрылымы, ондағы лексикалық материалдың берілу әдістәсілі (эквиваленция, дефиниция, безендірілу, лексика-семантикалық топтағы басқа сөздерге қарама-қарсылығы), синонимия, анонимия, паронимия, ассоциативтік, тіркесімдік, грамматикалық, орфоэпиялық және стильдік ерекшеліктері және т.б.

5) полиграфиялық орындалуы.

Лексикографияның жалпы теориясын дұрыс ескеру тілдік ақиқаттықты шынайы көрсете білуге мүмкіндік жасайды. Ал методикалық талаптар оқу сөздігіндегі ең маңызды деген жайттарды, мысалы, сөздіктің тілді үйренуге бағытталуын, белгілі оқыту кезеңі мен оқыту формасын қолдануын, яғни оның қызметін анықтайды.

Оқу лексикографиясының тәуелсіз лингводидактикалық пән болуының мәні, оның тіл білімі мен әдістеменің өзара тығыз байланыстылығы мен өзара қатыстылығында жатыр.

Шынында да, егер қазіргі түсінікте, оқу лексикографиясы тіл білімінің ең бір жас саласы бола тұра, практикалық тұрғыдан алғанда оны бұрыннан келе жатқан сала деуге де болады. Бұлай тұжырымдау үшін кез келген сөздікті, ең алдымен, дидактикалық шығарма деп түсінген жөн болар еді. Лексикограф, яғни сөздік авторы оқулықтың авторы тәріздес, қоғам (ұжымдық тіл иесі) мен жеке адам аралығындағы делдал ретінде қызмет етеді. Сондықтан кез келген сөздік типінің қызметі – жеке адамның санасындағы білім мен ұжым санасындағы білім аралығындағы айырмашылықты жоюға талпыныс жасайды деуге де болады.

Қолданбалы тіл білімінің ірі маманы В.Г. Гак сөздіктердің қызметін мынадай негізгі төрт топқа бөліп қарастырады:

1. Тіларалық қатынасты ұйымдастыруға байланысты. Мұндай қызмет көбінде екітілді сөздіктер мен тілдескіштер арқылы іске асады.

2. Тілді оқыту (үйрету) үшін және мағынасы қиын сөздерге түсініктеме беру арқылы.

Сөздіктің мұндай қызметінде лексикографиялық материалдар ретінде біртілді немесе екітілді сөздіктер, глоссариялар, тілдескіштер қолданылуы мүмкін. Оқыту процесі мынадай мәселелерге қатысты болады:

а) сойлеу тілінен айырымы бар көне мәдениет тіліне қатысты тілдерді меңгеруде. Мысалы, ол ежелгі дәуірге қатысты тіл немесе басқа да тілдерге – латын, ежелгі грек тілі, Еуропадағы ортағасырлық ескі словян тілі, Шығыс елдеріндегі классикалық араб тілі және т.б. тілдер болуы мүмкін;

ә) шет тілдеріне;

б) ана тілге.

3. Ана тілін сипаттау мен нормалауға байланысты. Мұндай қызметтер, ең алдымен, түсіндірме сөздіктер арқылы және орфографиялық, қиын сөздер сөздігі және т.б. арнаулы біртілдік сөздіктер арқылы іске асады.

4. Тілдерді зерттеу ісіне байланысты. Мұның негізінде біртілді және екітілді филологиялық сөздіктердің әртүрлі типтері құрастырылады (*Гак, 1977*).

Лексикографияның аталған қызметтерін тарихи тұрғыда қарастырсақ, оның даму сатыларын былай жіктеуге болар еді:

1. Сөздікке дейінгі кезең (Еуропада – XV ғ. дейінгі кезең). Бұл кезеңде сөздік нәтижелері глоссариялар, вокабулалар (хрестоматияға не оқулық тарауына қатысты берілетін қысқаша сөздік), ономастикондар түрінде көрініс тапты. Олардың негізгі мақсаты мәдениет тілінің түсінуге қиын сөздерінің мағынасын ашу.

2. Сөздік түзудің алғашқы кезеңі (Еуропада – XV–XVII ғасырлар). Бұл кезеңде лексикондар (белгілі қызмет аясына ғана тән сөз қоры) мен тіларалық қарым-қатынас және мәдениет тілін зерттеу мен шет тілін оқыту үшін қызмет ететін екітілдік сөздіктер өмірге келді.

3. Лексикографияның дамыған кезеңі (Еуропада – XVIII ғ. бастап). Бұл кезең сөздіктердің әртүрлі типтерінің, әсіресе, түсіндірме сөздіктердің шыға бастауымен ерекшеленеді (*Марусенко, 1996, 310*). Құрастырылған сөздіктер лексикография саласының жоғарыда аталып өткен төрт түрлі қызметін іске асыруды мақсат етеді.

Сонымен, оқу лексикографиясы сөздік құрастырудың алғашқы даму кезеңінің өзінде-ақ пайда болды деуге болады. Себебі, бастапқы кезеңдердегі сөздіктердің аттарының өзі-ақ, олардың дидактикалық міндеттерін нұсқап отырды. Мәселен, бірінші латынша-ағылшынша сөздік "Promptulorium parvulo-rum" деп аталған, яғни "Балаларға арналған сөз қойма". Ертеректегі французша-латынша сөздіктер – "Commentarius puerorum", яғни "Жастарға арналған түсіндірме сөздік" деп және т.б. аталған. Бері келе мұндай сөздіктердің оқуға бағыттағыш сипаты екінші орынға ауысып, біріншіге – түсіндірме, тарихи, этимологиялық сөздіктер типтеріне ауыса бастайды.

Лексиканы оқыту мақсатының нысаны етіп алуға арналған әдебиеттерді талдай келе, ғалым В.В.Морковкин мынандай негізгі бес аспектіні бөліп қарастырады (*Актуальн. проблемы ...*, 1977, 31-35 бб.):

1. Оқу сөздіктері. Мұндай сөздіктерге қатысты мәселелер, әсіресе, сөзтізбе құрылымын талдау мен оқу сөздіктерінің типтері жайлы мәселелер толық зерттелді деуге болады. Бірақ бұл зерттеулердің негізгі мақсаты әр типті сөздіктерде орын алатын ұқсастық пен айырмашылықтардың аналитикалық мәселелері, лексиканың семантикасын анықтаудың тәсілдері және тағы басқалар.

2. Лексикалық минимум сөздіктері. Қазіргі кезде тілдің сөздік қорын минимизациялау теориялық жағынан айтарлықтай зерттелді деуге болады. Соның нәтижесінде оқу формасына қарай көптеген лексикалық минимум сөздіктері құрастырылды. Бірақ оқу процесінде лексикалық минимумның барлық артықшылығын орынды пайдалану үшін ол жан-жақты сипатта жүзеге асуы және оқу барысының барлық кезеңдерін түгел қамтуы қажет. Әрине, бұл аталған мәселелердің әлі де толық шешімі табылмай келеді.

3. Лингвостатистикалық зерттеулер. Мұндай зерттеулердің нәтижелері, әдетте, жиілік сөздіктер мен үлестірімді жиілік сөздіктер (қолдану жиілігі әр стиль бойынша үлестіріліп берілген) түрлерінде көрініс табады.

4. Лексикографиялық типтегі лексикаға қатысты оқу құралдары. Әртүрлі тілдерге арналған мұндай оқу құралдары көптеп шығарылды деуге болады. Бірақ олар, теориялық тұрғыда, әлі де болса зерттеу нысаны бола алмай отыр.

5. Оқулықтар. Мұндай оқулықтарда лексиканың семантикалық жақтары әр тақырыпқа қатысты пәндік сөздіктерде, жинақтық (сводный) сөздіктерде және арнайы лексикалық жаттығуларда қарастырылады. Қандайда болсын оқулықтар үшін лексикалық аспект өте-мөте маңызды деп саналады. Бірақ мұндай оқулықтарда лексиканы семантикалау тәжірибесінің теориялық негіздемесі орын ала бермейді.

Жоғарыда көрсетілгендей, оқу сөздігі, ең алдымен, қажетті әдістемелік бағытталуымен, нақты оқу кезеңдеріне қатыстылығымен, оқудың мақсаты мен міндетіне және т.б. қарай сипатталады. Оқу сөздіктері тілдік материалды таңдау ісінде арнайы мақсат қоюмен ерекшеленеді. Яғни болашақ лексикографиялық талдау үшін қажет болатын тірек бірліктердің негізгі тізімін түзу ісімен ерекшеленеді. Сөздіктің сөзтізбесі әрі шағын, әрі жеткілікті сипатта болуы қажет. Бұл мақсат тек сөздің негізгі мағынасындағы квантитативті параметрлерін, грамматикалық тұрпатын және синтаксистік құрылымын, тақырыпты ашатын ең маңызды сөздер мен сөз тіркестерді таңдауды қатаң түрде есепке алу арқылы ғана жүзеге асады. Осымен бірге сөздіктің негізгі тақырыбы мен сөйлеу жағдаятын қамтуды тексеру қажеттігі де туындайды. Өкінішке қарай, тәжірибе жүзінде мұндай тексеру жұмысы көбінде орын ала бермейді.

Оқу сөздігінің сөзтізбе көлемі оның белгілі бір типіне қарай ғана анықталады. Мысалы, қысқаша түсіндірме сөздіктегі лексика осы тілдік ұжымның әр мүшесі үшін қалыпқа айналған сипатта болуы қажет. Аталған мәселені шешу үшін мәдениеттің және ғылым мен өндірістік білімнің орта жиынтығы қамтылған ең кіші сөйлеу қызметінің моделі құрастырылуы керек. Мұндай түсіндірме сөздік-анықтағыш 20000–25000 сөздік мақаладан тұратыны белгілі. Бірақ мұндай сөздік пассивті немесе жеке индивидтердің активті сөздік қорының шекаралығында тұрады да, оқу сөздігінің активті типіне жатады.

Кез келген оқу сөздігі қатаң мақсаттылықпен жүргізілген таңдаудың нәтижесі, яғни кең түсініктегі лексикалық минимум бола тұра, ол ішкі диалектикалық қарама-қарсылыққа толы. Бір жағынан, ондай сөздік нақты бағыттағы мәтіндерді және оқыту кезеңіне қатысты кейбір қиындықтарды да қамтуы қажет. Ал екіншіден, ондай сөздіктерде кез келген мәтінді түсінуге мүмкіндік туғызатын нақты материалдар да орын алуы керек. Бір ғана принципті (жиіліктік, тақырыптық, ситуативтік және т.б.) ұстаным ететін сөздік типі қойылған талапты орындай алмайды.

Сөздіктің әр деңгейіне тән сөздер бірнеше топ құрайды. Оларға жататындар (*Марусенко, 1996, 313*):

а) тілдің грамматикалық қаңқасын құрайтын және кез келген мәтінде кең қолданыс табатын – құрылымдық сөздер;

ә) көптеген мәтіндерде қолданыс табатын және сол себепті нақты тақырыпқа, сөйленіс жағдаятына байланыссыз келетін – жиілігі жоғары негіз сөздер;

б) белгілі тақырыппен немесе жағдаятпен тығыз байланысты болатын – тақырыптық, жағдаяттық сөздер. Бұл категорияға қатысты сөздер жоғары жиілікке ие емес, шектеулі сөйлеу мәнмәтіндерінде қолданылады. Сөздікке тек әртүрлі категориядағы (құрылымдық, тақырыптық, жағдаяттық) бірліктерді енгізу ғана кең шеңбердегі мәтіндерді түсінуге мүмкіндік туғызады.

Оқу сөздігінің құрылымдық бөліктерінің өзгешелігі. "Үлкен" (ірі) сөздіктерге қарағанда оқу сөздіктерінің құрылымы негізгі екі ерекшелікпен сипатталады: лексикалық бірліктерге түсініктеме берумен (семантикалау) және ол бірліктердің парадигматикалық, синтагматикалық қасиеттері мен қатынастарын анықтайтын әдіс-тәсілдерімен.

Лексикалық бірліктерді семантикалау түсіндірме оқу сөздігінің маңызды құрылымды бөлігі болып табылады. Әрине мұндай компонент белгілі дәрежеде басқа да сөздіктер типіне де қатысты. Аударма оқу сөздіктерінде семантикалау,

әдетте, аударма баламалары арқылы, ал түсіндірме сөздіктерде мұндай семантикалау – анықтамалар арқылы жүзеге асады. Оқу сөздіктерінің маңызды құрылымдық бөлігі саналатын анықтама: шағындық, көрнекілік және қарапайым сөзге саятындай талаптарға сай келетіндей болуы керек.

Шетел азаматтарына арналған түсіндірме сөздіктерде қайсыбір ұғымға түсініктеме беретін анықтамаларында таныс сөздердің шағын көлемді бөлігін пайдаланған жөн болады. Бұл сөздер "минимум сөздік" құрайды және осылардың негізінде сөздіктің басқа да лексикалық бірліктері анықталады. Мысалы, кең танымал М.Уэстің (*West, 1965*) сөздігінде 24000 сөздің мағынасы көлемі 1500 сөздерге тең лексикалық топ мағыналары арқылы анықталады. Сол сияқты француз тілінің негізгі лексикалық тобын қамтитын (*Gougenheim, 1959*) анықтағышта 55 сөз арқылы және қарапайым сөздіктің 1374 сөздері бойынша басқа сөздердің мағыналық анықтамалары берілген.

Оқу сөздіктерін құрастыру тәжірибесінен байқалатын жайт, ол "минимум сөздіктер" негізінде түзілген сөздіктер, көбінде, жасандылық пен нақтылықтың жеткіліксіздік сипатымен өзгешеленеді. Бірақ мұндай олқылықтар семантикалау әдісінің кемшілігінен емес, құрастырушылардың жіберген қателігінен деуге болады. Оқу сөздіктері үшін аса маңызды мәселе сөздердің синтагматикалық және парадигматикалық байланысын бейнелеп көрсете білу. Үйренуге тиісті тілдің тек лексикасын ғана меңгеру сөйлеу тіліндегі оның дұрыс қолданылуына кепіл бола алмайтындығы белгілі. Ана тілі мен шет тілінің түйісу әсерінен үйренушінің тілінде бұрын-соңды қолданылмайтын кейбір сөз тіркестерін туындатады. Ірі филологиялық сөздіктерде сөзқолданыстардың ерекшеліктері жайлы, яғни сөздердің тіркесімдігі жайлы мәліметтер бола бермейді. Сондықтан тәжірибе жүзінде, осындай мәліметтер бере алатын, сөздердің ең ықтимал нормативті байланысын бейнелейтін оқу сөздіктері қажет-ақ (*Регинина и др., 1980*).

"Парадигма" дегеніміз бір-біріне қарама-қарсы қойылған, бірақ ортақ белгісі бар немесе ұқсас ассоциация тудырушы кез келген лингвистикалық бірліктер тобы екені белгілі. Міне, осындай парадигматикалық қатынастарды бейнелейтін сөздіктерде тақырыптық, синонимдік, антонимдік топтар көрініс табады.

Оқу сөздіктерінің типологиясы. Типологиялық тұрғыда оқу сөздіктері бірнеше көзқараста қарастырылады:

1) сөздіктің қызметі тұрғысынан;

2) сөздіктің экстралингвистикалық параметрі бойынша;

3) сөздіктің ішкі лингвистикалық параметрі бойынша.

Оқу сөздіктері қызметіне қарай: **жалпылама** және **мамандандырылған** сөздіктер деп те ажыратылады. Бірінші сөздік үйренушілердің барлық категорияларына арналады да, ал екіншісі олардың тек белгілі категорияларына ғана арналады. Мысалы, мектеп оқушыларына, филолог-студенттерге, филолог емес студенттерге, мұғалімдерге және т.б.

Сөздіктің экстралингвистикалық параметрлері әрбір сөздіктің өзіне тән ерекшелігін анықтайды. Мысалы, олардың энциклопедиялық сөздікке жақындығына қарай немесе оның кейбір элементтерінің ғана энциклопедиялық сөздікке енгендігі бойынша. Осы тұрғыда мынадай бірліктерді бөліп қарастыруға болады: материалдық мәдениеттің заттық сипаты, рухани мәдениеттің концептері, ел тану компоненті және үйренуге (зерттеуге) қажетті тілдегі логикалық категориялардың бейнелену сипаты. Яғни "әлем бейнесінің" әртүрлі жекелік, топтық, ұлттық санасының концептуалды бейнеленуі. Экстралингвистикалық параметрлер жалқы есімдердің, мақал-мәтелдердің, қанатты сөздердің (сөйлемшелер) және т.б. сөздіктері үшін аса маңызды рөл атқарады.

Ішкі лингвистикалық параметрлер негізіндегі сөздіктер типологиясы тілдің сөздік құрамының типологиясын бейнелеуі қажет. Кез келген тілдің сөздік құрамы жүздеп-мыңдап саналады. Сондықтан қажетті сөзді тез іздеп табу үшін лексиканың жүйелік сипатын ескеретін стратегияны туындатып алған жөн болады. Соның негізінде қажетті деген сөзді барлық сөздік құрамының ішінен іздемей, тек оның кішігірім бөлігінің ішінен, мысалы, семантикалық өріс ішінен немесе лексика-семантикалық топ ішінен және т.б. іздеу мүмкіндігі туындайды. Оқу лексикографиясының теориясы, ең ұтымды жағдайда, тіл жүйесінің лексикалық аспекті бойынша, аспектілік сөздіктер жүйесін құруды қалайды. Сол арқылы іздеудің табиғи стратегиясы модельденіп, қажетті сөзді іздеп-табу мезгілі қысқарады.

Бұл мәселе сөздіктің көлеміне, оларды ұйымдастыруға, сөздіктегі әрбір лексикалық бірлікті сақтайтын ақпарат көлеміне, әрқилы таңбалар жүйесіне, сілтемелер жүйесін құру жұмыстарына қатысы бар. Яғни бұл мәселе кешенді (жан-жақты) сипатқа ие деуге болады.

Қазіргі кезде халықаралық қатынас нығайып, көптеген дербес ұлттық мемлекеттердің пайда болуы және бұрынғы отарлықта қалған халықтардың мәдениетінің көтерілу сипаты, қайтадан лексикографияның тіларалық қарым-қатынасқа және тілдерді оқытуға байланысты қызметтерін бірінші орынға ауыстыруына жағдай жасайды. Әлемдік өндірісті интеграциялау кезеңінде әртүрлі түсініктерді бірізге түсіру ісі де аса қажет болуда. Сондықтан егемен Қазақстанның халықаралық деңгейге шығуына және бірқатар жетекші елдермен қазіргі таңдағы ғылым мен технологияның әртүрлі салаларында ғылыми және өндірістік ақпараттармен алмасудың едәуір ұлғаюына байланысты неғұрлым кең ауқымды қамтитын көптілді сөздіктер шығару қажеттілігі де туындауда. Жақын болашақтағы лексикографиялық зерттеулердің негізінде оқу сөздіктерінің әр ұлттық тілдерге, соның ішінде қазақ тіліне қатысты да практикалық маңызы зор жүйесі құрылуы мүмкін.

Сонымен, оқу лексикографиясының ғылыми пән ретіндегі ерекшелігі жайында жоғарыда сөз болған мәселелер және де осы бағытқа қатысты басқа да теориялық жайттар ескерілетін болса, қазақ тілінің оқу лексикографиясына байланысты көптеген әрекеттерді компьютер көмегімен толық автоматтандырып, олардың электронды түрлерін көптеп шығару жүзеге асар еді.

Актуальные проблемы учебной лексикографии. М., 1977.

Гак В.Г. О некоторых закономерностях развития лексикографии (учебная и общая лексикография в историческом аспекте) // Актуальные проблемы учебный лексикографии. М., 1977.

Жантекеева З.Ө., Нақысбеков О. Қазақша-орысша және орысша-қазақша оқушыларға, студенттерге арналған сөздік. Астана, 2006. 208 б.

Жұбанов Б.А. Химиялық сөздік ағылшын, неміс, орыс және қазақ тілдерінде. Алматы: Ғылым, 1988. 200 б.

Исмағұлова Б., Ережепова Э., Әбдіжапарова Г. Оқушыларға, студенттерге арналған қазақша-орысша, орысша-қазақша сөздік. Алматы: Аруна, 2002. 416 б.

Марусенко М.А. Учебная лексикография // Прикладное языкознание: Учебник / Отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. 528 с.

Момынов Т., Рақышев А. Медициналық сөздік (орысша-қазақша). Алматы: Атамұра, 1999. 312 б.

Регинина К.Б., Тюрина Г.П., Широкова Л.И. Устойчивые словосочетания русского языка. М., 1980.

Салқынбай А., Абақан Е. Лингвистикалық түсіндірме сөздік. Алматы: Сөздік-Словарь, 1998. 304 б.

Gougenheim G. Dictionnaire fondamemtal de la langue francaise. Paris, 1959.

West M. An International Reader's Dictionary. London, 1965.

ҒЫЛЫМИ-ТЕХНИКАЛЫҚ ЛЕКСИКОГРАФИЯ ЖӘНЕ АРНАУЛЫ СӨЗДІКТЕР ТҮРЛЕРІ

Қазақ тіліндегі ғылыми-техникалық стиль және оған тән терминдік жүйе жаңадан ғана қалыптасып келеді деуге болады. Себебі техникалық ғылымдардың қазақ тіліндегі көрінісі немесе дерек көздері түрлі монографиялық зерттеулер, оқулықтар, мақалалар, баяндамалар, тезистер, патенттер т.б. сан-салалы еңбектер арқылы танылады. Ал бұл аталғандардың саны мен сапасы қазақ ғалымдарының әлі де болса арнайы көңіл бөлуін қажет етеді. Ана тілімізде жарық көре бастаған жаңа бастаманың нәтижесінде тілімізге жаңа ұғымдар мен техникалық жаңа атаулар келе бастағаны белгілі. Сондықтан қазақ тілінің жалпы лексикография саласының жаңа бағыты ретінде ғылыми-техникалық лексикография саласы да өз отауын құра бастады деуге болады.

Үнді-еуропа тілдерінде, соның ішінде орыс тілі бойынша аталған бағыттың өзіндік даму тарихы мен теориялық және практикалық тәжірибесі баршылық. Міне, осындай жайттарды ескере отырып, осы еңбекте ғылыми-техникалық терминологиялық сөздіктерге немесе арнаулы сөздер сөздігін құрастыруға қойылатын кейбір талаптарға қысқаша тоқталмақпыз.

Арнаулы сөздіктерді құрастырудың теориясы мен практикасы ғылыми-техникалық лексикографияның зерттеу нысаны болып саналып жүр. Мұндай арнаулы сөздіктер арқылы баспа материалдары мен әртүрлі мекеме орындарының іскағаздарында көрініс тапқан терминдер мен әр ғылымдар салаларында орын алатын номенклатуралық деректер жайлы түсініктеме мен анықтамалар алуға болады.

Арнаулы сөздіктер белгілі бір сала маманы мен электронды машиналар қажеттіктері үшін құрастырылады. Ғылыми сала маманына қатыстылар: біртілді және аударма терминологиялық сөздіктер, идентификаторлар сөздігі, терминдердің жиілік сөздіктері, рубрикаторлар мен классификаторлар сөздіктері. Арнаулы сөздіктердің екінші тобына, яғни электронды машиналардың қажеттіктеріне – түрлі типті ақпаратты жүйелерге арналған ақпараттық-іздестіру тезаурустары (*Тезаурус...*) және мәтінді автоматты өңдеу жүйелерінде қолданыс табатын машиналық сөздіктердің (*Убин*) әртүрлі типтері (негіз сөздердің, тіркестердің, қосымшалардың) жатады.

Кез келген тілдің дәстүрлі түсіндірме сөздіктеріне қарағанда ғылыми-техникалық терминологиялық сөздіктер, ең алдымен, белгілі білім аясы мен техникалық саланың мамандарына арналатыны белгілі. Терминологиялық сөздіктер, өздерінің бағышталуына байланысты – жоғары білімді мамандар мен ғылыми әдебиеттерді аударумен айналысатын кәсіптік аудармашыларға арналып құрастырылады. Жоғары дәрежелі мамандарға мұндай сөздіктер тек қажетті деген сөзді тауып алуға ғана мүмкіндік туғызып қоймай, сонымен бірге оның ұғымын анықтауға және әртүрлі басылымдарда кездесетін кейбір синонимдік сөздер мағыналарын ажыратуға да мүмкіндік туғызады.

Ғылыми мақалаларды бір тілден екінші тілге аудару кезінде идеографикалық принципте құрылған біртілді және көптілді түсіндірмелі-аударма терминологиялық сөздіктер де аса қажет (*Берков; Убин*).

Ғылыми-техникалық терминологиялық сөздіктерді орта және жоғары техникалық оқу орындарының студенттеріне, әр салалық өндіріс орындары мен

ауылшаруашылық аясындағы сан-салалы мамандарына және радио мен баспа қызметкерлеріне және т.б. арналған сөздіктер деп те ажыратуға болады. Сол сияқты, мектеп оқушыларына арналған терминологиялық сөздіктер де жеке пішінде шығарылып жүр. Бұл сөздіктердегі деректерді оқырманға ұсыну тәсілінде оқушының жас ерекшелігіне айрықша көңіл бөлу негізгі ұстаным болып табылады. Мысалы, терминологиялық сөздікте қызықтыру элементтерімен бірге суреттердің, сызбалардың орын алуы. Соңғы кездерде ғылыми-көпшіліктік терминологиялық сөздіктермен қатар бөлек ғылыми аяның терминдік жүйесін анықтайтын арнаулы ғылыми сөздіктер де жарық көре бастады. Әдетте, мұндай сөздіктердің құрылымы әліпби тәртібінен гөрі, жүйелілік принципке негізделіп жасалады.

Қолданылу аясына қарай, тіркейтін, түгелдейтін – "узус" сөздіктерін, белгілі ережелерге бағындыруды көздейтін – регламенттеуші сөздіктерін, оқу жұмысын көздейтін – мәлімет беру, жүйелендіруші терминологиялық сөздіктер де жарық көруде (*Герд, 1973; 1996*).

Тіркеу міндетін атқаратын немесе "узус" атты сөздіктер қатарына терминологиялық сөздіктердің басым көпшілігінің қатысы бар. Ол сөздіктер сол кезеңге және қарастырылып отырған білім саласына тән терминологиялық стандартты анықтайды.

Регламенттеуші сөздіктер: **нормативті** және **нормативті емес** деп екіге ажыратылады. Нормативті терминологиялық сөздіктерге айқын түрде мысал бола алатын сөздіктер – стандарт терминдер сөздігі мен ұсынылған (бекітілген) терминдердің сөздігі.

Стандарт терминдердің анықтамалары бойынша олар – ресми, мемлекеттік-құқықты статусқа ие болатын жүйелі типтегі терминологиялық сөздіктер. Терминологиялық стандарттардың құрамы – кіріспе, негізгі және әліпбилік көрсеткіштерінің бөліктерінен тұрады.

Кіріспе бөлікте стандарттың таратылу аясы, стандарттау нысанын анықтау және қолданылуға болмайтын терминдер, синонимдер мен аударма баламалары жайлы мәліметтер беріледі.

Стандарттың негізгі бөлімінде осы стандартқа қатысты қабылданған ұғымдар жүйесіне сәйкес орналасқан терминдер мен олардың анықтамалары қамтылады.

Терминологиялық сөздіктерге әр білім саласы бойынша Ғылыми-техникалық терминологиялық комитетінің жүйелі түрде шығарып тұратын ұсынылған (бекітілген) терминдердің жинағын да жатқызуға болады. Мұндай жинақтарда негізгі терминдермен бірге олардың анықтамалары, синонимдері және аударма баламалары да беріледі.

Нормативті терминологиялық сөздіктерді құрастыру аса маңызды іс. Сондықтан ондай сөздік бір ғана терминдік жүйеге қатысты болуымен бірге олардағы терминдердің қолдану шарттылығы арнайы құжат арқылы бекітілуі қажет.

Арнаулы ғылым-техникалық терминологиялық сөздіктерді құрастырудың алға қойған мақсатына қарай, олар – **ақпараттаушы** және **жүйелендіруші** деп екіге ажыратылады.

Арнаулы ақпараттаушы сөздікке мысал ретінде ақпараттық-іздестірудің тезаурустық түрін атауға болады. Ақпараттық-іздестіру тезаурусы деп біртілді қадағаланып отыратын және лексикалық бірліктерге өзгеріс енгізіп тұру сипатындағы сөздікті атайды. Бұл сөздік бір ғана тілдің лексикасына негізделіп, ақпаратты өңдеу мен іздестіру қажеттігі үшін құрастырылады. Ақпараттық-іздестіру тезаурусы, ол – ақпаратты автоматты өңдеу мен іздестіруді жеңілдететін сөздік түрі. Ондай тезаурус терминдер арасындағы мағыналық, семантикалық қатынасты

да толық ашуға жәрдем етеді. Ақпараттық-іздестіру тезаурусның негізгі бірлігі – терминдер мен дескрипторлар. Оның әліпбиленген лексика-семантикалық бөлігі – бір жерге жинақталған дескрипторлық сөздік мақалалары. Әрбір дескрипторлық сөздік мақала термин-дескриптордан бастама алады да, ал одан кейін дескриптордың мақала шеңберіндегі синонимдері беріледі және негізгі дескриптормен грамматикалық немесе ассоциативті қатынаста болатын басқа да терминдер орналасады (*Тезаурус ...).*

Жүйелендіруші арнаулы сөздіктерге жататындар: *классификаторлар, рубрикаторлар, жиілікті сөздіктер* және *терминдік жүйе* сөздіктері. Классификаторларда реттелген түрде ғылым мен техниканың жүздеген нысан атаулары және олардың сипаттамалары беріледі. Мұнда ең алдымен техника-экономикалық көрсеткіштердің классификаторларын, өндіріс пен ауылшаруашылық өнімдер атауларын және т.б. атауға болады. Осындай терминологиялық сөздіктер ішінен әсіресе рубрикатор-сөздіктер маңызды рөл атқарады. Рубрикатор-сөздіктерді сала мамандары мен жеке мекемелер үшін ақпараттық қызмет ету жүйесіндегі әдістемелік құжат деп те атауға болады.

Рубрикаторлар, көбінде, өндірістік саланың негізгі бағыттарының жүйеленген ретін және маңызды нысандар мен процестерді қамтиды.

Салалық жиілік сөздіктердің өзіне тән маңыздылығы – кейбір бөлектелген терминдердің жиіліктерінің қамтылуы. Сөздіктің ең бір аса маңызды сипаты – терминдердің тақырыптық сипатта қамтылуы. Егер жиілік сөздік бірнеше пәндік аяны қамтитын болса, онда ол көпсалалық сөздік деп аталады. Көбіне, ғылымның дербес бөлініп шығуынан тарсалалық сөздіктер жасау қажеттігі туындайды. Мұндай сөздіктерді жасау тәжірибесінде түсіндірме терминологиялық сөздік типтерін құрастыру да аса маңызды. Түсіндірме терминологиялық сөздіктердің міндеті – нақты бір сала бойынша анықтама құралы болу және барлық терминдерді қамтумен бірге ұғымдар жүйесінің көрінісін беру. Мұндай мәліметтер мамандардың ғылыми және тәжірибелік жұмыстары үшін аса қажет деп саналады.

Терминологиялық түсіндірме сөздікте ғылым мен техниканың дамуынан пайда болатын барлық жаңа ұғымдардың терминдік көрінісін берумен қатар, ол ұғымдардың анықтамалары да қамтылуы қажет.

Терминологиялық сөздіктердегі мәліметтерді реттеуде дәстүрлі әліпбилік және ұялы әдістермен бірге статистикалық әдіс те қолданыс табады. Яғни мәліметтерді реттеуде терминнің мәтіндегі жиі немесе сирек кездесу сипаты да ескеріледі (*Алексеев*).

Әртүрлі тілдердің деректерін қамту жағынан терминологиялық сөздіктер біртілді және екітілді, үштілді және көптілді аударма сөздіктер деп бөлінеді (*Берков; Убин*). Түсіндірме сөздіктерді жасаудағы негізгі ұстаным – терминнің жаңа мағынаға ие болуы.

Терминологиялық сөздіктердің әртүрлі типтерінің негізгі дереккөзі ретінде жетекші ғалымдар мен маман-тәжірибешілердің қаламынан жарық көрген белгілі бір білім саласына қатысты жинақтап-қорытындылаушы монографиялар мен мақалалар материалдары алынады. Сонымен бірге қажетті материалдардың қатарына жоғары оқу орындарының ірі оқулықтарын да жатқызуға болады. Әдетте, тек осындай оқулықтарда ғана ғылыми ұғымдар мен терминдердің аса айқын және нақты анықтамалары беріледі. Энциклопедиялар мен басқа да сөздіктерде берілген салалық терминдердің барлығы бірдей ғылыми-техникалық терминологиялық сөздіктерде орын алуы қажет.

Ғылыми-техникалық терминологиялық сөздіктердің лексикалық қорына қосымша ретінде рефективті журналдардағы кітап мазмұнының рефераттары,

жаңалықтары, патенттер мен ісқағаздарының деректері де қамтылып алынады. Аса "қысылған" және шоғырланған түрде берілген мұндай мәліметтер негізгі терминдер жайлы жеткілікті түрде түсініктеме бере алады. Сол сияқты, Ресейдегі ғылыми-техникалық әдебиеттер мен құжаттарды аудару орталығының жүйелі түрде шығарып тұратын "Тетради новых терминов" атты еңбектің материалдары да терминологиялық сөздіктер үшін құнды дереккөз болып табылады.

Неше түрлі құрал-саймандарды жөндеуге қатысты және т.б. салалық, ведомстволық нұсқаулар мен нұсқаулықтарда кездесетін терминдер мен номенклатуралық белгілеулердің терминологиялық сөздік жасаудағы атқаратын рөлі де айтарлықтай мәнді.

Сонымен, сөздік түзу кезіндегі әртүрлі дереккөздерден таңдалып алынған терминдер негізінде терминологиялық сөзтізбе құрастырылады. Бұндай сөзтізбелердің реестрлік бірлігіне жататындар: сөз, сөз тіркес, қысқартулар, символдар және олардың тіркестері. Әдетте терминологиялық сөздіктің сөзтізбесі таза интуитивті түрғыда түзіледі, яғни әрбір сала мамандарының жеке көзқарасындағы негізгі және маңызды деп аталатын терминдері таңдалып алынады. Әрине мұндай әдіс ғылыми әдебиеттерді талдауға негізделмеген, тек қана субъективті пікірлік ұстаным екені белгілі.

Ғылыми-техникалық терминологиялық сөздікті құрастырудан бұрын, маман-лексикограф ең алдымен баспаға дайындалып жатқан еңбектің хронологиялық шеңберін анықтап алуы қажет. Ғылымның жаңа салалары, мысалы, есептеу техникасы, радиофизика, биофизика үшін мұндай мәселе онша өзекті болмауы да мүмкін, себебі, ондай сөздікке енетін кез келген лексиканың уақыт мерзімі беріден басталады. Ал жаратылыстану ғылымдарының – физика, механика, математика, химия, биология, география, геология және көптеген қоғамдық ғылымдар салаларында мәселе басқаша. Бұл ғылымдардың орысша мәтіндері өзінің тарихи дамуында 150–200 жылдай мезгілді қамтиды. Осы себептерден терминологиялық сөздікте көрініс табатын терминдердің хронологиялық тәртібі маңызды мәселелер қатарына жатады.

Аталған мәселені тарихи-филологиялық және тарихи-ғылыми тұрғыда кең түрде қарастыратын болсақ, ол ертеректегі мәтіндерді дұрыстап оқи білу, оның мағынасын дұрыс түсіну және бір тілден екінші тілге дұрыс аудару мәселелеріне қатысты деуге болады. Ежелгі терминдердің де қазіргі әдебиеттерде қолданыс табатынын ескерсек, аталған мәселе әр уақытта да өз құндылығын жоймайтыны белгілі.

Кейбір ғылыми кітаптар мен мақалаларда өткен кезеңдердегі ғалымдардың зерттеу нәтижелерімен пікірталас болып жатады. Мысалы, XIX–XX ғасырлардағы ғалымдардың зерттеулеріндегі ұғымдар мен концепцияларына көңіл бөліну жағы, көбінде, филология, математика, биология, химия салаларына қатысты болады. Егер мұның бәрі тек қана табиғи тіл арқылы ғана іске асатынын ескерсек, терминологиялық сөздіктерге лексикалық қорды таңдап алу кезеңдік сипатта болатынына және тіл туралы ғылым тарихының өзі де ғылым тарихымен ұштасып жататынына көз жеткізуге болады.

Мәселен, орыс тіліндегі ғылыми-биологиялық мәтіндердің тарихы мынадай кезеңдерге бөлінеді:

1) XVIII ғ. бастап, XIX ғ. 50-жылдар аралығы – орыс тіліндегі биологиялық мәтіндердің бастапқы қалыптасу кезеңі;

2) XIX ғ. 50–60-шы жылдары мен XX ғ. бастапқы кезеңі (10-20-шы жылдар);

3) XX ғ. 10–20-шы жылдары мен бүгінгі күнге дейінгі кезең.

Биология саласы бойынша терминологиялық сөздік құру барысында XX ғ. алғашқы жылдарына дейінгі биологияға қатысты терминдерді қазіргі биология тіліне жатқызуға болмайды (*Герд, 1986; 1996*).

Әрине, терминологиялық сөздікті пайдалану кезінде сол сөздік ішінен термин жайлы, тіпті оның тарихына дейінгі мәліметтер табылып жатса, әртүрлі сөздіктермен әрекет етудің қажеттігі болмас еді. Әрі тарихи тұрғыда терең хронологиялық қабатты қамтып, әрі қазіргі кездің терминдері бойынша анықтағыш рөлін атқаруға қатысты төмендегідей екі түрлі маңызды ой-пікір бар.

Біріншіден, ол – ғылыми білім саласының айтарлықтай күрделі және әрдайым құбылып тұру сипаты, ұғымдар иерархиясының күрделі эволюциясы, ұғымдардың мазмұны мен көлемінің үнемі өзгеру сипаты, ұғымдар арасындағы терминдердің тұрақсыз үлестірілуі, терминологиядағы синонимдік қатарлардың тұрақсыздығы.

Екіншіден, басқа тілдердің лексикалық параллельдерін ескере отырып, оқырманға қазіргі терминология жайлы молынан мәлімет беру жағдаяты. Сондықтан терминологияның даму тарихын қамтитын өзінше бір бөлек арнаулы сөздік болғаны жөн.

Қалай болғанда да терминологиялық сөздіктің хронологиялық төменгі шегі жайлы мәселе әлі де болса өз шешімін таппай тұр. Төменгі хронологиялық шек XX ғ. 20–30-шы жылдары ма немесе XIX ғ. мен XX ғ. аралығы ма? Оны тек нақты білім және ғылым салаларының негізгі кезеңдеріндегі тілдің даму сипатына қарай ғана шешуге болады (*Герд, 1973; 1996*).

Терминологиялық сөздіктердің сөзтізбесін қалыптастырғанда субъективті пікірді азайтудың бірден-бір жолы – нақты ғылым мен білім саласына тән логика-ұғымдық жүйені алдын ала модельдеу. Арнайы мәтіндерді логика-ұғымдық тұрғыда талдау тәсілі терминологиялық сөздіктерді жасаудағы алдын ала қойылатын қажетті шарт деуге болады.

Терминологиялық сөздіктердің құрылымы жайлы мынадай екі негізгі мәселені атауға болады:

1) терминологиялық сөздіктегі мәліметтерді қалыптастыру принциптері жайлы мәселе;

2) терминологиялық сөздіктегі жеке фрагменттердің құрылымы жайлы мәселе.

Бірінші мәселеге байланысты мынаны айтуға болады. Сөздік материалын ұйымдастыруға қатысты ғылыми-техникалық лексикографияда классикалық лексикографиядағы сияқты әліпбилік, әліпбилік-ұялы, идеографиялық жалпы принциптер қарастырылады.

Терминологиялық сөздіктердің басым көпшілігі әліпби принципімен құрастырылатыны белгілі. Бұл ұстаным кез келген сөздікті пайдалану мен ондағы бірліктердің орналасу тәртібі жағынан көптен бері өзін-өзі ақтап келеді.

Осы айтылғандарға қарамай, мамандардың көптеген практикалық мұқтаждығынан аңғаратынымыз, ол терминологиялық сөздіктердегі терминдердің орналасуы тек әліпбилік немесе әліпбилік-ұялы принциптерге ғана сүйенбей, сонымен бірге логика-ұғымдық негізде, яғни ұғымдардың қатал ғылыми классификациясы бойынша идеографиялық принципте жасалуының қажеттігі. Бұл жердегі идеографикалық принципті қайсыбір ұғымды таңба арқылы белгілеп жазу деп түсіну керек. Мұндай принцип нақты ғылымның, білім саласының даму қисынына сәйкес келуімен бірге, терминді мағынасы бойынша сөздіктен тез тауып алуға да мүмкіндік жасайды. Мәселен, нақты ғылымдар мамандарына, көбінде, қайсыбір ұғымның қалай және қандай сөзбен аталатыны немесе ол сөздің синонимдерінің қандай ұғымдармен байланысы бар екендігін білу қажет болады.

Маманның зерттеу тәжірибесінде және аударма практикасында белгілі бір терминді тұрпаты жағынан ұқсас болып келетін ертеректердегі зерттеулердің терминдерімен сәйкестендіру қажеттігі де туындауы мүмкін (*Герд, 1996).*

Идеографикалық принциппен құрастырылған терминологиялық сөздіктің негізгі құрамы – сол жағындағы терминдердің әліпби тәртібімен берілуі ғана емес, ол вербальды түрде берілген және шартты немесе формалданған тұрпаттағы ұғым болуы қажет. Аталған ұғым сөздіктің жалпы логика-ұғымдық құрылымында қатаң түрде өз орнына ие болады. Мысалы, Роженің тезаурусында (*Roget's thesaurus..., 1978*) сөздік мақаланың басы ұғымның коды арқылы (үш орынды сан) және осы ұғымды білдіретін тірек сөзбен басталады. Тірек сөзден кейін сызықша және сөз табының белгісі қойылады, сосын сол сөздің синонимдері беріледі.

Доризайфтың неміс тілінің идеографиялық сөздігінде (*Dorn Sejf, 1958, S. 211*) әрбір сөздік мақала белгілі бір тақырыптық топтағы ұғымның нөмірімен басталып, әрі қарай құрылымдық түрғыда бір-біріне жақын жатқан осы ұғымды білдіретін сөздің синонимдік баламалары берілген.

Ғылыми-техникалық терминологиялық сөздіктің жеке фрагменттерінің құрылымы. Лексикографияның жалпы теориясына сәйкес, сөздік мақала, көп жағдайда, мынадай бөліктерден тұрады (*Герд, 1996, 297-280*):

1) сөздік мақаланың аты;

2) терминдердің грамматикалық сипаттамасы;

3) терминнің қолданысына байланысты стильдік реңкі мен пайда болуына қатысты сипаттамасы;

4) терминнің семантикалық сипаттамасы.

Әліпбилік принципте түзілген терминологиялық сөздік мақаланың құрылымы түсіндірме және энциклопедиялық сөздіктердің құрылымына ұқсас болып келеді. Терминологиялық сөздіктерде сөздік мақаланың аты, әдетте, әліпби тәртібімен орналасады да, олар тек термин сөздер ғана емес, сөз тіркестер, сөздер мен символдардың комбинациясы, аббревиатуралық қысқартулар болуы мүмкін.

Әліпбилік принциппен құрылған ғылыми-техникалық терминологиялық сөздіктерде терминнің семантикалық сипаттамасын беру оның мағынасының логикалық анықтамасы арқылы немесе аударма терминологиялық сөздіктердегі баламасы арқылы іске асады. Терминнің әрі шағын, әрі дәл логикалық анықтамасының берілуі, оның семантикалық құрылымын сипаттаудың ең бір тиімді жолы деп саналады.

Терминологиялық сөздіктерде терминдердің мағынасын анықтаудың ең бір дұрыс жолы, ол әрбір ұғымның типіне қарай, біріздікті сақтай отыра, алдын ала белгілі параметрлерге сүйену. Терминнің толық түрдегі лексикографикалық анықтамасын құруға қатысты мынадай талаптарды ұстаным еткен жөн: логика-ұғымдық сызбаға алдын ала талдау жасау, ұғымның шығу тегін анықтау, ұғымға тән маңызды белгілерді анықтау, осы ұғымның басқа ұғымдар аясы ішінен ерекше белгілерін анықтау.

Арнайы әдебиетті аудару ісімен айналысатын аудармашыға қажетті сөздің баламасын табумен қатар оның синонимдерін де іздеп табу қажет болады. Яғни терминді мазмұндық түрғыда кеңірек танып, сөздікте көрсетілген сөз, шындығында да түпнұсқадағы ұғымды білдіру-білдірмеуіне көз жеткізу керек болады *(Берков; Убин).*

Ғылыми-техникалық терминологиялық сөздіктерді құрастыруда терминнің анықтамасын тұжырымды етіп беру үшін мынадай ережелерді басшылыққа алған жөн саналады *(Герд, 1996, 298-300):*

1. Термин арқылы аталатын барлық ұғымдар мен нысандар анықтама ішінде қамтылуы қажет және, керісінше, анықтамада сөз болатын барлық ұғымдар мен нысандар терминнің атындағы ұғымдар мен нысандарға сәйкес келуі қажет.

2. Терминнің анықтамасында пайдаланылатын белгілер қарастырып отырған **тақырыптық ая** бойынша ең маңызды және мәнді болуы қажет.

3. Терминнің анықтамасында басқа бір терминнің анықтамасына сілтеме жасауға болмайды. Себебі, егер ол терминнің анықтамасында тағы да басқа бір терминнің анықтамасына жүгінсе, оның шегі табылмай, қайта оралып алғашқы терминге келуіміз де мүмкін.

4. Анықтамада пайдаланылған терминдер осы сөздік ішінде анықталған болуы немесе бұрыннан да белгілі және бірізді түсінігі болуы шарт. Егер анықтамада пайдаланылған терминдер іргелес жатқан басқа пәндерге қатысы болса немесе ол жалпы-пәндік термин болса, онда арнайы қосымшаларға сілтеме жасалғаны жөн. Басқаша жағдайда ол терминнің түсінігін қандай әдебиеттен оқып-білуге болатыны айтылуы керек.

5. Терминнің нақты мағынасын ашатын анықтама берілуі үшін оның ғылыми ұғымдар жүйесіндегі алатын орнын жете түсіну аса маңызды.

6. Терминдердің анықтамалары, көптеген жағдайда, сөздермен қатар, математикалық және химиялық формулалар мен неше түрлі безендірілулер арқылы көрініс табады.

7. Әр елдің тілдеріндегі бір ғана терминнің мазмұнына қатысты анықтамалар бір-біріне сәйкес болмауы мүмкін. Сондықтан терминнің анықтамасын берерде басқа бір сөздікке сүйенетін болсақ, оның қай елдің сөздігінен алынғаны көрсетілуі керек.

Біртілді немесе аударма терминологиялық сөздіктерде терминнің семантикалық сипаттамасы синонимия, антонимия, текті-түрлік және ассоциативтік семантикалық қатынастар арқылы толықтырылуына әрі кеңейтілуіне болады.

Ғылыми-техникалық терминологиялық сөздікте нақты білім, ғылым, техника салаларының терминдік жүйесі көрініс табуы қажет. Ғылыми мәтіннің семантикалық мазмұнын дәл ашу, сол сала маманының қатысуынсыз мүмкін емес екендігі белгілі. Сондықтан ұғымның түсінігін сипаттау сол ғылыми сала маманының ісі де, ал оны барынша дәл және ықшам түрде тұжырымдап оқырманға жеткізу тілші-лексикографтың үлесі. Ал ғылыми-техникалық терминологиялық сөздіктің сапасы сала маманы мен тілші-лексикограф маманының бірлесе әрекет жасауының нәтижесі деуге болады.

Жоғарыда аталған арнаулы сөздіктерді құрастырудың теориясы мен практикасына байланысты тұжырымдар қазақ тілінің ғылыми-техникалық лексикографиясының дамуына оң әсерін тигізеді деген ойдамыз.

Алексеев П.М. Статистическая лексикография. М., 1975.

Берков В.П. Вопросы двуязычной лексикографии. М., 1973.

Герд А.С. Основы научно-технической лексикографии. Л., 1986.

Герд А.С. Научно-техническая лексикография // Прикладное языкознание: Учебник / Отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. С. 287-307.

Тезаурус научно-технических терминов. М., 1972.

Убин И.И. Автоматический переводной словарь. М., 1989.

ЛЕКСИКОГРАФИЯ САЛАСЫНДАҒЫ АВТОМАТТАНДЫРУ ТӘСІЛІ

Лексикографиялық нысанның ерекше түрі ретіндегі *автоматты сөздіктер* мен *автоматты терминологиялық банктердің* пайда болуы бірнеше себептермен түсіндіріледі. Біріншіден, ол ЭЕМ-ның мүмкіндіктерінің артуы болса, екіншіден, лексикографтар мен аудармашылардың еңбек өнімділігін арттыру жолдарын іздестіру қажеттігін ұғыну деуге болады.

Лексикография саласы адамның қызметтерінің ішіндегі ең бір көп еңбек сіңіруді керек ететін түріне жататыны мәлім. Сөздіктердің кейбір түрін түзу жұмысында оны сапалы аяқтап шығу үшін бірнеше онжылдықтардың өтуі мүмкін. Сондықтан лингвистикалық жұмыстарды автоматтандырудың алғашқы жетістіктерінің лексикография саласына қатысты болуы кездейсоқ жағдай емес. Неше түрлі сөзтізбелерді, үлкен көлемді сөзнұсқағыштарды, жиілік сөздіктер мен кері-әліпбилі сөздіктерді мақсатқа сай өмірге келтіру үшін оларды ЭЕМ көмегімен түзудің қажеттігі қазіргі кезде дау туғызбайды.

Автоматты сөздіктердің негізгі міндеті – адамның мәтінмен әрекет етуінде оның еңбек өнімділігін арттыру. Мысалы, лексикографиялық түсініктеме беру мақсатындағы әртүрлі тіл бірліктерін жинақтау, сақтау және өңдеу жұмыстарын автоматты сөздіктер арқылы іске асырған жөн. Яғни сөздіктерді олардың функционалдық критерийі негізінде қарастыратын болсақ, автоматты сөздіктерді айрықша сөздік түрі деуге болады. Оның міндеті – мәтінді автоматты өңдеудің практикалық жүйесінде қолданыс табу.

Кез келген автоматты сөздіктердің міндетті түрдегі бөлігі ретінде сөзтізбе алынады. Мұндай сөзтізбе семантикалық баяндау жүйесіне немесе білім базасына енудің тетігі болумен бірге тілдік базаның бірден-бір құрастырушысы да бола алады. Мысалы, ақпараттық қызмет ету жүйелерінде түзілетін сөздіктер өз құрамында тиісті білім базасына (тезаурус, фрейм жүйелеріне) ену тетігі рөлін атқаратын дискрипторлық сөзтізбелерді сақтай алады. Ал машиналық аударма жүйелеріндегі сөзтізбелер мәтінге лексикалық және синтаксистік талдауды іске асыру үшін қажетті деген морфологиялық, синтаксистік және семантикалық сипаттамаларға ие лексикалық бірліктерді өз құрамында қорландырады. Осымен бірге мұндай жүйелер әртүрлі мағыналарды, синтаксистік және семантикалық ережелерді сипаттайтын автоматты сөздіктерді де өз құрамына енгізулері қажет. Мұндай автоматты сөздіктер машиналық аударма жүйесінің білім базасын құрайды.

Кез келген тәжірибелік жүйе жұмысы да өз құрамында білім базасына жол салатын сөзтізбе болуын қалайды.

Сөйлеу тіліндегі ақпаратты өз жадына ендіруге бейімделген жүйелердегі сөзтізбе құрамында сөзтұлғалардың жазылуы мен оның транскрипциясы да орын алады.

Түсіндірме типті біртілді автоматты сөздік сөзтізбесі сөздік жобасының принциптеріне сүйене отырып анықталады. *Автоматты аударма сөздіктері* негізінен ғылыми-техникалық мәтіндермен әрекет етуге арналған болғандықтан, олардың көпшілігі терминологиялық лексиканы қамтулары тиіс. Ірі автоматты аударма сөздіктің нәтижелігін арттыру үшін оған тақырыптық сөздікке қатысты барлық терминологиялық лексика қамтылуы қажет. Яғни жақын тақырыптық аядағы аса жиі қолданыстағы жалпы лексика мен жалпығылымдық, жалпытехникалық

лексиканың барлығы бірдей есепке алынады. Жалпы-лексикалық қор, ең алдымен, техника саласындағы мамандарға қажеттірек екені белгілі. Себебі, ондай мамандар, әдетте, өз саласына қатысты лексиканы жақсы меңгергенімен, шет тіліне қатысты лексиканы қажетті дәрежеде біле бермейді.

Автоматты сөздіктің міндеті ғылыми-техникалық мәтіндердің барлық түрлерімен әрекет жасау болғандықтан, оның құрамына барлық терминдер түрлері енулері қажет. Мысалы, оларға жататындар: стандарттанған және ұсынылған, сол сияқты ұсынылмаған, тұрақтанған немесе жаңадан қолданысқа енген терминдер.

Автоматты сөздіктің құрылымын анықтау жолында мынадай негізгі үш мәселенің шешімі табылғаны жөн:

1) сөздік мақаланың басқы бірлігінің тұрпатын анықтау;

2) тілдік бірлікті сипаттауға қатысты ақпарат жиынтығын анықтау;

3) автоматты сөздік базасындағы сөздік мақаланы құрастырудың ұтымды тәсілін таңдау.

Автоматты сөздік үшін оның сөзтізбесінің негізгі бірлігін таңдау маңызды мәселе болып саналады. Жалпы лексиканы қамтитын ірі сөздіктерде сөзтізбенің негізгі бірлігі – **негіз сөз** ретінде алынады. Ал әрбір сөзге қатысты сөз тіркестері, әдетте, мақала ішіндегі мағыналық көрініс ретінде немесе басқы тілдік бірліктердің тіркесімдік сипатына мысал ретінде беріледі.

Автоматты аударма сөздікке қатысты сөзтізбенің негізгі бірлігі дәстүрлі терминологиялық сөздіктердің осындай бірліктеріне жақын сипатта болулары қажет. Сөзтізбедегі лексикалық топтың 80 пайызы сөз тіркестерден тұратыны да ескерілуі қажет. Ең ірі автоматты аударма сөздік, көпшілік жағдайда, дәстүрлі терминологиялық сөздіктерге жүктелетін мәселелерді шешуге арналған терминологиялық банк рөлін атқарады. Мұндай автоматты аударма сөздіктерде дәстүрлі терминологиялық сөздіктерге қарағанда сөз тіркестерінің жоғарғы шегі анық болмауынан, оларда сөз тіркестер саны тұрақсыз болып келеді. Ірі автоматты аударма сөздіктерде дербес мақала ретінде жеке сөйлемдер мен мәтін үзінділері де есепке алынады.

Аударма сөздіктердегі сөзтізбені іріктеу мен жинақтаудың аталған түрлері, әдетте, дәстүрлі лексикографияда қолданыла бермейді. Ал автоматты сөздіктер үшін мұндай тәсіл белгілі жағдайда ұтымды болып келеді. Себебі стандартты мәтіндер бойынша жартылай автоматтанған аударма жасау тұстарында автоматты аударма сөздіктер өздерін тиімді құрал ретінде көрсете білді. Мысалы, заң құжаттары мәтіндерінде белгілі бір мәтін бөліктері онша өзгеріссіз қайталанып отыратыны белгілі.

Қазіргі кездегі өндірістік автоматты аударма сөздіктер аса күшті ЭЕМ негізінде құрылады. Мұндай ЭЕМ бірнеше жүзмыңдаған көлемді сөз тіркестерінен тұратын тілдік ақпараттармен нәтижелі әрекет ете алады. Сонда да ЭЕМ әрекетінің жылдамдығы оның жадының көлеміне тікелей байланысты екендігін ескере отырып, оны үнемді пайдалануды әр уақытта ескеру қажет.

Екітілді және көптілді лексикография жағдайында сөздік бірлігін таңдау нақты тілдердің сөзжасам және тұлғажасам ерекшеліктерімен анықталады.

Айырғыш, аналитикалық, кейбір жағдайда флективті-аналитикалық тілдер типтерін өңдеуге қатысты тез әрекет ететін жүйелер құруда "*Мәтін – Автоматты сөздік*" өзара әрекеттесу сызбасы өзін-өзі ақтап жүргенін айта кету қажет. Мұндай жағдайда сөздіктегі белгілі бір сөзге тіркелген барлық ақпаратпен танысу үшін синтаксистік деңгейге дейінгі морфологиялық талдау алгоритмінің көмегіне сүйенбеуге де болады. Сонымен бірге автоматты сөздіктегі сөзтұлғаларды олардың

мәтіндегі "егіздерімен" теңдестіру аса үлкен жылдамдықпен іске асады. Ал синтетикалық, флективтік және агглюнативтік тілдердің автоматты сөздіктерінің бас бірлігі ретінде сөзтұлға алынатын болса, сөздік көлемі кенет ұлғаяды да мәтін бойындағы сөзтұлғаны автоматты сөздіктегі сөзтұлғамен теңдестіру жылдамдығы, керісінше, азаяды. Мысалы, орыс тіліндегі зат есім сөз 12-ге дейін, ал етістік сөз, егер түр корреляциясын есепке алатын болсақ, 188 рет тұлға өзгерісіне ұшырайды екен. Түркі тілдерінде есім түріндегі тілдік бірлік саны, теориялық жағынан алғанда, тұлға өзгерісіне шексіздікке дейін ұшырауы мүмкін.

Осы айтылғандарға байланысты мұндай парадигмаларды компрессиялау (қысу) қажеттігі туындайды.

Бұл мәселенің шешімі екі түрлі тәсілмен іске асуы мүмкін:

1) лингвистикалық (алгоритмдік);

2) техникалық.

Лингвистикалық тәсілде негізгі назар сөздіктің бірлігін қалыптастыруға бөлінеді және оның эталонын, үлгісін және ең ыңғайлы баламасын табуға әрекет етіледі. Синтетикалық, флективтік және агглюнативтік тілдер үшін неше түрлі компрессиялау тәсілдері қолданылып жүр. Олардың ішінен ең нәтижелі деген әдіс ретінде автоматты сөздікті негіз сөздердің сөздігі ретінде қалыптастыру болып табылады. Мұндай автоматты сөздіктерді, әдетте, тілдік бірліктің екі түрі жайлы сөз етуге болады: *негіз сөз* және *ерекше тұлға*. Басқаша айтқанда олар сөз негізінің алмасуы арқылы немесе супплетивті (әртүрлі негізден жасалған бір сөздің тұлғалары) тәсіл арқылы жасалған сөзтұлғалар. Ерекше тұлға бірліктер автоматты сөздіктің құрамына негіз сөздермен бірдей дәрежеде енгізіледі.

Сонымен бірге бірқатар тілдер үшін парадигма ретінде тек жалғыз негіз сөзді ғана енгізу жеткілікті бола бермейді. Бұл жағдай тек ерекше тұлғаларды қолдану кезінде де жеткіліксіз.

Дамыған флексиясы бар орыс тілінің лексикалық жүйесін модельдеу тұрғысынан негіз сөздер сөздігі қолданылады. Мысалы, А.И.Герцен атындағы РМПУ-дың инженерлік зертханасы зерттеп дайындаған және СИЛОД жүйесінің өзегі болып саналатын орыс тілінің "МАРС" автоматты сөздігі негіз сөздер мен типтік парадигмалар сөздігі негізінде құрылған. Ал GETA жүйесін сөздікпен жабдықтау үшін орыс тілінің мынадай жеті сөздігін іске қосу алдын ала қарастырылған: алғашқы екеуі – екі негіз сөздер сөздігі, келесі екеуі – өзгеріске ұшырамайтын сөздер мен ерекше тұлғалардың, тағы да екеуі – жұрнақ пен жалғаулардың сөздігі және соңғы бір сөздік – машиналық тіркесім сөздігі.

Сөздік мақаланың құрамы әртүрлі автоматты аударма сөздіктерде өзгеше және олардың зоналық айырмашылығы 10-нан 99-ға дейін құбылып отырады. Аталған зоналарда әр типті лингвистикалық және экстралингвистикалық ақпараттар тіркеледі.

Әр автоматты аударма сөздік өзінің сөздік мақаласының ішінде белгілі реттілікпен мынадай мәліметтерді қамтиды: басқы сөз, оның бір немесе бірнеше шеттілдік баламалары, басқы сөздің тақырыптық қатыстылығы және оның шеттілдік баламалары, грамматикалық ақпарат, формалданбаған түсініктемелер немесе стандарттанған дефинициялар, мәнмәтіндер, лексикалық тіркесім, басқы сөздің семантикалық иерархиялық байласымы, сапалық коды (сенімдік индексі), қысқартулар, стильдік мәнері (бояюы), лингвогеографиялық шектелуі, тірек сөздер (егер басқы лексикалық бірлік сөз тіркес болса), дерек көздер, сөздік мақаланың авторы, шығу мезгілі, әртүрлі қызметтік ақпараттар.

Сөздік мақаланың құрылымындағы семантикалық, синтаксистік және морфологиялық қатынастар арқылы мақаланың өзіне тән жазылу ерекшелігі анықталады.

Деректер базасында орын алатын әрбір лексикалық бірлік морфологиялық, синтаксистік, семантикалық және функционалдық деңгейлерде сипаттамалар жиынтығы ретінде көрініс табады. Синтаксистік және морфологиялық сипаттамалар мәтін бойында кездесетін сөзтұлғаның мағынасын анықтау үшін және оның сөйлемдегі синтаксистік қызметін белгілеу мақсатында қажет. Семантикалық сипаттамалар лексикалық бірліктің мағыналық жағын баян етеді де, ал функционалдық сипаттамалар бірліктің барлық құжат құрылымындағы потенциалды рөлін анықтайды.

Қазіргі кездегі мәтінді автоматты өңдеу жүйелерінде лексикалық бірліктің жоғарыдағы сипаттамаларына тиісті орын беру мақсатында сөздік мақалалардың төмендегідей үш түрі қолданыс табады.

1. Көлемі алдын ала белгіленген, бірыңғай қатал құрылымды сөздік мақала. Бұл құрылым лексикалық бірлік тобын сипаттайтын барлық өлшемдерді мақала ішінде қамту әрекетін алдын ала қарастырады. Мұндай тәсіл лексикалық бірліктерді баяндаудың барлығына бірдей қатысты. Оған лексикалық бірліктің семантикалық және функционалдық өлшемдері де қоса беріледі. Аталған тәсіл бойынша лексикалық бірлікті баяндау тек синтаксистік ақпаратпен ғана шектелуі мүмкін. Бұл жағдайда лингвистикалық ақпараттық базаға семантикалық ақпарат жайлы мәлімет жеке сөздік ретінде енгізіледі.

Осындай мақала құрылымында ақпараттың барлық типтері мен олардың **байтық** көлемдік шегі алдын ала анықталып, белгіленеді. Егер ақпараттың нақты типі анықталмаса, оның орны "бос" күйінде сақталуы қажет. Мұндай қатал құрылымды сөздік мақалаға мысал ретінде жалпы көлемі 225 байт алатын SYSTRAN жүйесіндегі мақаланы атауға болады. Бұл жүйенің автоматты сөздігі негіз сөздер сөздігі ретінде құрылған. Әрбір сөздік мақала өз бойында мынадай мәліметтерді сақтайды:

Y – лингвистикалық ақпараттық базаға енгізу кезіндегі лексикалық бірлікті баяндау тәсілі (жаңадан қосу, түзету, жою);

A – базалық машиналық негізді баяндау (описание);

B – септік жалғаулар жайлы морфологиялық ақпарат, көптік жалғаулардың түр-тұрпаты және т.б.;

C – меңгеруге қатысты мәліметі бар синтаксистік кодтар;

Д – семантикалық кодтар;

E – сөз алдында келетін көмекші сөздердің кодтары (барлық сөз таптары үшін бұл орын бос болуы керек);

F – балама ерекшеліктерін баяндауды қамтитын негізгі аудармалар.

Сөздік мақаланы құрудың мұндай әдісі ондағы ақпаратты оңай жолмен толықтыру мен үнемі өзгертіп отыруға барынша мүмкіндік тудырады. Себебі автоматты сөздікте әрбір ақпараттың алдын ала белгіленген өз орны болғанымен, оларда ең басында ескерілмеген жаңаша типті ақпаратты базаға енгізу сөздік мақаланы, сөздікті және базаның өзін түбегейлі өзгертуге әкеп соғады. Сондықтан мәтінді автоматты өңдеу жүйелерінде лексикалық бірліктің барлық сипаттамалары толық қамтылғаны жөн болады. Әрине мұндай талап, әсіресе өндірістік жүйелерде ЭЕМ-ның қомақты жадын алуымен бірге автоматты сөздікті орналастырудың қиындық тұстарына әкеп соғады. Солай бола тұра бұл әдіс, көбінде, ақпаратты екінші рет қайталаудан сақтандырады.

2. Оң жақ шегі жылжымалы болып келетін сөздік мақала. Мұндай мақаланың баяндалуы – "таңба – мағына" принципі бойынша жүзеге асады. Аталған тәсіл бойынша мақаланы баяндауды қалыптастыру бірізді сызба арқылы жүргізіледі және лексикалық бірлікке қатысты барлық ақпаратты зоналарға үлестіру жолдары

қарастырылады. Ондай зоналардың саны нақты мәліметтердің типіне байланысты болып келеді. Ал лингвистикалық бірлікті баяндау үшін, сөзді топтастыру тұрғысынан алғанда, лексика-грамматикалық, морфологиялық, семантика-синтаксистік, тұрпаттық, функционалдық және т.б. көзқарастағы сипаттамаларға сүйенеді. Әрбір зонада лексикалық бірліктің қандай топқа қатыстылығына қарай, белгілер саны жекеленіп алынады. Мысалы, зат есімді актуалдаудың функциясын атқаратын детерминативтер класы осы кластың нақты бірлігін баяндайтын мынадай белгілер жиынтығымен сипатталады: сөздің ішкі топтары, өзгеріп түрленуі, орналасу қалпы және т.б.

Сөздік мақаланың "оң жақ шегінің жылжымалы" болуы мақаланы тек ЭЕМ-ға (компьютерге) дайындау кезіне ғана сәйкес келеді. Ал компьютердің жадына енгізу тұсында мақаланың баяндалуы, алғашқыдай қатаң құрылымға ауыстырылуы керек.

3. Иерархиялық құрылымдағы сөздік мақала. Бұл тәсіл ақпарат бөліктерін жеке файлдар (массивтер) ретінде бөліктеп шығаруды алдын ала көздейді. Мұндай жағдайда автоматты сөздік, ең аз дегенде, екіге ажыратылады: міндетті түрдегі тіркелген ақпараты бар сөздікке ену тізімі және мағыналар сөздігі.

Лексикалық бірліктің мәтінді автоматты өңдеу жүйесінде қолданылу мүмкіндігін ескере отырып, оның барлық сипаттамаларын иерархиялық топтастыру алдын ала жүргізіледі, яғни иерархиялық жоғарғы және төменгі деңгейлері анықталады. Мәтінді автоматты өңдеудің барлық түрлерінде тікелей қолданылатын жоғарғы деңгейдегі сипаттамалар мақала құрылымына тікелей енгізіледі, ал барлық басқа сипаттамалар (әдетте семантикалық) жеке файл түрінде көрініс табады. Мұндай жағдайда сөздік мақалада жекешеленген файлдарға арнайы сілтеме беріледі.

Базалық вариантта сөздік мақала құрылымының (базаға "енетін") сөздік құрамына жататындар: басқы лексикалық бірлік (негіз, сөзтұлға немесе сөз тіркесі); ендірілетін тілдің типологиялық ерекшелігіне, грамматикасына байланысты лексика-синтаксистік код; сөздің әртүрлі заттық аяда атқаратын қызметінің ерекшелігін анықтайтын ішкі тіл (подъязык) коды; аударма жайлы ақпарат.

Баяндаудың жоғарғы деңгейі, яғни сөзтізбе өз құрамында лексикалық бірлікті және сөйлемге алғашқы синтаксистік (синтагмалық) талдау жүргізуге қажетті морфологиялық сипаттамаларды қамтиды. Мысалы, сұрақ-жауап жүйесінде баяндаудың жоғарғы деңгейі – әрқайсысына синтаксистік және морфологиялық ақпарат тіркелген лексикалық бірліктің өзінен ғана тұруы мүмкін. Бұл деңгей **анықтағыш** деп аталып, іштей екі бөліктен тұрады (*Беляева, Герд, Убин, 324-с).*

Енді автоматты аударма сөздік арқылы орындалатын лингвистикалық баяндаулардың кейбір ерекшеліктеріне тоқталайық.

Бұл мәліметтер Бүкілодақтық аударма орталығында (ВЦП) зерттеліп, сөздік мақаланың құрылымына енгізуді көздейді. Сөздік мақала 13 зонаға бөлініп, әрбір зона қатаң түрде белгілі ақпаратпен өзіндік ереже бойынша толтырылады.

1-зона. **Сөздік мақаланың атын анықтайтын тірек зона.** Бұл зона бойынша сөздік мақалаға "ену" әрекеті жүзеге асады және оны қалыптастыру мен іздестіру жұмыстары орындалады. Мақаланың аты ретінде: морфема, жеке сөз, канондық тұрпаттағы сөз тіркесі немесе мәтін үзіндісі алынуы мүмкін. Толық омонимдер үшін рет саны тіркелген әртүрлі сөздік мақалалары құрылады.

2-зона. Сөздік мақаланың атынан кейін оның **лексика-грамматикалық класын** анықтайтын 2-зона орналасады. Бір элементті лексикалық бірліктер сөз таптарына қатыстылығына қарай топ-топқа бөлінеді. Бұл топтардың ішінде лексикалық бірліктер бірнеше категориялық белгілері бойынша тағы да ішкі кішігірім топтарға

ажыратылады. Сөз тіркестері өздерінің сөйлем ішіндегі атқаратын қызметіне қарай типтік түрлерге бөлінеді.

2а-зона. **Орыс тіліндегі аударма эквивалентінің лексика-грамматикалық класын анықтайтын зона.** Тікелей орыс тіліндегі аударма эквивалентінен кейін оның лексика-грамматикалық класы жазылады. Орыс тіліндегі аударма эквивалентінің сөздік мақаланың құрамына енуі сөздік мақаланы "керісінше" орын ауыстыру үшін қажет, яғни *шет тілі – орысша* сөздік мақаладан автоматты түрде, керісінше, *орысша – шет тілі* сөздігінің мақалаға ауыстыру мақсаты үшін.

3-зона. **Морфо-грамматикалық ақпарат зонасы.** Бұл зонада басқы бірліктің барлық сөз өзгерту парадигмаларын құру әрекеті орындалады да, соның нәтижесінде кез келген мәтіндік түрпатты іздестіру мүмкіндігі туады.

Морфо-грамматикалық ақпарат зонасында басқы лексикалық бірліктерді қысқартып жазу варианттары ескерілетін ішкі зона қарастырылған. Мұндай мүмкіндік жеке тұрған сөздерге де, сөз тіркестеріне де қатысы бар.

3а-зона. **Орыс тіліндегі аударма эквивалентіндегі қысқартулар зонасы.** 2а-зонадан кейін 3а-зона енгізіледі. Яғни лексика-грамматикалық кластан соң орыс тіліндегі аударма эквивалентінің қысқартулар зонасы орналасады.

4-зона. **Тақырыптық рубрика (сөзбасы) мен стиль зонасы.** Бұл зонада тақырыптық және стильдік белгілер беріледі. Ондай белгілер көмегімен басқы бірліктің осы мағынадағы қолдану аясы мен шарты нақтыланады. Осы зонада шет тілдерінен алынған кірме лексиканың дереккөзіне сілтеме жасалуы мүмкін. Осындай белгілер мен таңбалардың типтері негізінде белгілі стилистикалық ерекшелігі бар монотематикалық сөздіктерді немесе сөзтізбелерді автоматты түрде қалыптастыруға болады.

5-зона. **Сенімділік индексінің зонасы.** Әрбір орыс тіліндегі аударма эквивалентінің алдында осы аударма эквивалентінің (баламасының) көпшілік мақұлдаған сенімділік индексі тұрады. Бұл индекс лексикалық бірлік алынған лексикографиялық дереккөздің сенімділігімен анықталады. Сенімділік дәрежесін белгілеу үшін орыс тіліндегі аударма эквиваленті жүйесінде мынадай индекстер қабылданған:

А – қайсыбір стандартта (ГОСТ, ОСТ, халықаралық стандарт) бекітілген немесе оның қолданысын беделді халықаралық мекеме ұсынған аударма эквиваленті (баламасы).

Б – аударма эквиваленті сенімді және апробациядан өткен сөздіктерде берілген.

В – аударма эквиваленті жаңа терминдер дәптерінде, терминологияның экспресс-шығарымдарында және т.б. берілген.

Г – аударма эквиваленті аударылған журналдарда, баспадан шықпаған аудармаларда кездескен немесе сөздік мақаланы құрастырушының ұйғарымымен алынған.

6-зона. **Орыс тіліндегі аударма эквивалент зонасы.** Бұл зона канондық түрпатта жазылған сөз немесе сөз тіркестерінен тұрады. Барлық орыс тіліндегі аударма эквиваленттер қатар сандарымен беріледі. Синонимдік қатарды құрайтын орыс тіліндегі аударма эквиваленттері бірдей нөмірмен жазылады, ал басқы бірліктің әртүрлі мағыналарын білдіретін лексикалық бірліктің аударма баламалары әртүрлі нөмірленеді.

Дәстүрлі лексикографияда *баламасыз лексика* деген ұғым бар. Мұндай лексикаға басқа тілдерге аударылмайтын лингвистикалық бірлікті жатқызады. Баламасыз лексика басқа елдерде жоқ ұғымды білдіретін жалпы қолданыстағы лексика немесе басқа тілдерде әлі қолданыла қоймаған жаңа терминдік лексика болуы мүмкін.

Егер жаңа термин өзінің аударма баламасының мағынасын ашып жатса, оны орыс тіліндегі аударма эквивалентіне қосып жазуға болады. Бұдан кейін сөздік мақала бір қалыпқа түседі.

7-зона. **Түсіндіргіш белгілер (таңбалар) зонасы.** Түсіндіргіш белгілер немесе таңбалар лексикалық бірліктердің мағынасын нақтылау үшін және олардың мәтіндегі қолдану ережелерін сипаттау үшін қажет. Бұлардың ақпараттағы аударма баламасын дұрыс таңдап алуға да жәрдемі бар. Сөздік мақалаларын "кері қарай аудару" жағдайында, яғни **орысша-шетелдік** нұсқаны автоматты түрде қалыптастыру жағдайында түсіндіргіш таңбалар шетел лексикалық бірлік рөліндегі аударма баламасында да өз орнын сақтайды. Түсіндіргіш таңбалар лексикалық бірлікке қойылатын талаптарға сай, семантикалық терминдер негізінде де көрініс табуы мүмкін. Яғни орыс лексемасына сәйкес келетін бас актанттың орнында қандай семантикалық топтың өкілі тұратынына нұсқайды.

Семантикалық таңбалар басқы лексикалық бірліктің және оның аударма баламасының семантикалық сипатының да көрінісі болуы мүмкін.

Лексикалық таңбалар орыс тіліндегі аударма эквивалентті және басқы лексикалық бірліктің лексикалық тіркесімдік жағдайын нақты лексемалар тұрғысынан анықтай алады.

Синтаксистік таңбалар орыс тіліндегі аударма эквиваленттерді басқарудың синтаксистік моделін баяндайды. Бұл жағдай синтаксистік орын мен басқы бірліктің осы мағынадағы грамматикалық әрекетінің ерекшеліктері тұрғысынан сипатталады.

8-зона. **Түсінік беру зонасы.** Терминологиялық бірліктердің мағынасын баяндау үшін түсініктеме беру тәсілін қолдануға болады. Автоматты аудару сөздігі жүйесінде омонимдік және көпмағыналық құбылыстардың шекарасын ажыратуда түсініктеме беру ең бір ұтымды тәсіл болып табылады.

9-зона. **Басқы бірліктің қолдану жағдаятына мысал келтіру зонасы немесе тілдік бірліктің қолдану мәнмәтіндерінің зонасы.** Басқы лексикалық бірліктің белгілі мағынадағы қолдану мүмкіндіктері орыс тіліндегі аударма эквивалентінде тиісті мысалдармен қамтамасыз етілуі мүмкін.

10-зона. **Фразеологиялық зона.** Басқы сөзді немесе сөз тіркесін, олардың лексикалық жүйеден алатын орны тұрғысынан сипаттайтын ақпаратты жазуға арналған ішкі зона (подзона) бөлініп алынады.

Мәтінде әртүрлі көрініс табатын лингвистикалық бірліктер жиі кездесетіні белгілі. Мұндай жағдайда лингвистикалық бірліктердің жазылу түрінің санына қарай, сөздік мақала саны құрастырылады.

11-зона. **Сөздік мақаланы құрастырушының немесе редактордың аты-жөнін тіркейтін зона:** фамилиясы, аты және әкесінің аты.

12-зона. **Жазба материалдың немесе пайдаланылған деректердің аттарын тіркейтін зона.** Әрбір аударма баламасына сөздік мақалада оның қандай дереккөзден алынғаны көрсетіледі.

13-зона. **Сөздік мақаланы құрастыру немесе редакциялау мезгілін тіркейтін зона**: жылы, айы, күні.

Жоғарыда сипатталған 13 зонаға қатысты нақты мысалдармен Л.Н.Беляева, А.С.Герд, И.И.Убиндердің "Автоматизация в лексикографии" атты мақаласы бойынша толығырақ танысуға болады (*Прик. язык, с. 318-333*).

Сөйлем түсінігінің синтаксистік, семантикалық көпмағыналығымен "күресу" үшін және нөлдік деңгейдегі семантиканы ұсыну қажеттігінен көбінде **сөз орамдарының автоматты сөздігі** қолданылады. Мұндай сөздік бойынша мәтіндегі

сөздер тізбектерін бөліп алып, оларды екінші тілге аударуға мүмкіндік туындайды. Сөздер тізбегінің мағынасы әрбір элементтің мағыналарының жиынтығы еместігі ескеріледі.

Лингвистикалық ақпараттар базасы құрамына аталған сөз орамдарының автоматты сөздігі дараланып немесе автоматты сөздікпен бірге енуі мүмкін. Бұл сөздіктердің (сөз орамы сөздігінің) қажеттігі қазіргі кездегі техникалық құралдардағы тілдік деректердің жеке сөздермен емес, сөз тіркестерімен берілетіндігімен түсіндіріледі.

Тіларалық идиомалық сөздердің машиналық аударма кезінде мәтін мағынасын тануға қиындық туғызатыны белгілі. Бұл қиындықты семантикалық талдау алгоритмін қолданып шешу де оңай емес. Сондықтан идиомаларды дұрыс аудару үшін сөз орамдары сөздігін машиналық аударма жүйесіне енгізу қажеттігі туындайды.

Сөз орамдарының сөздігін құрастыруда сегменттер шекарасын бөліп алу үшін статистикалық критерийлер ескеріледі. Мысалы, сөз орамдарының автоматты сөздікке ғылыми-техникалық мәтіндер бойынша жасалған әртүрлі жиілік сөздіктердегі аса жиі қолданылатын сөз тіркестері енгізіледі. Сонымен бірге ол сөздікке сөз тіркестерінің аударылу критерийлері де қоса беріледі. Басқаша айтқанда, сөз орамдарының автоматты сөздігінде бастапқы мәтін үзіндісінің тікелей аудармасымен сәйкес келе бермейтін мәтін үзінділері де қамтылады.

Тілдің семантикалық жағын баянды ететін автоматты сөздік дербес модуль ретінде бірнеше тәсілдермен құрылады:

а) белгілі ережелер негізінде біріктірілген семантикалық белгілердің жиынтығы түрінде;

ә) кеңейтілген немесе ассоциацияланған (байласымды) семантикалық желілер ретінде.

Әдетте, күрделі автоматты аударма жүйелерінде бірнеше тақырыптық аяның лексикалық топтары қамтылуы керек. Себебі әр тақырыпқа қатысты лексиканың бір-бірімен қиылысу мәселесінің шешімін табу автоматты аударма жүйесінде аса маңызды іс болып саналады. Өндірістік автоматты аударма жүйесін құрастырушылар аталған мәселеге үлкен мән беріп, осы мақсаттың шешімін табуға арнаулы лексиканы тақырып бойынша топтастыру әрекетін жүзеге асыратын жүйелерді іске қосуда. Автоматты аударма жүйесіндегі лексиканы тақырыптық топтауға негіз ретінде ғылыми-техникалық ақпараттың Мемлекеттік автоматтандырылған жүйесі орталығының басқы сөздері (рубрикаторлары) алынған.

Машиналық аударманың автоматты сөздігінің қатаң түрде нақты тақырыптық аяға бағышталуы лексикалық талдау деңгейінде тілдік бірліктердің көпмағыналығын ажыратуға және терминдерді аудару мәселесін стандарттауға мүмкіндік тудырады. Нақты тақырыптық аяға бағышталу автоматты сөздіктің кез келген жүйесі үшін бірден-бір қажет. Ал егер автоматты сөздіктің көпаспектілі жүйесін құру керек болса, онда тілдік бірліктің нақты тақырыптық аяға қатыстылығын ажырату қажеттігінен сөздік мақаланың құрылымына арнайы шартты белгі-код енгізіледі. Сол код арқылы лексикалық талдау кезінде тілдік бірліктің аударма баламасы таңдалады.

Көптілдік автоматты аударма жүйесін құрастырудың негізгі қағидасы ретінде автоматты сөздіктегі әр тілге қатысты тілдік бірліктердің өзара байласымдық сипатын анықтау тәсілі алынады. Өндірістік автоматты аударма жүйесінде тілдердің санына қатыссыз лексикааралық байланысты анықтаудың негізгі үш түрлі тәсілі қолданылады:

1. Әрбір лексикалық бірлік үшін басқа тілдерге қатыссыз, тек нақты осы тілдің өз мүмкіндігі (құралдары) негізінде ғана баяндалады (басқы сөз жайлы мәліметтер, грамматикалық ақпарат, лингвистикалық және т.б. да ақпараттар);

2. Аударма баламасы ретінде жүретін екі түрлі тілдердің тілдік бірліктеріне ортақ болып келетін сөздік мақала құрастырылады.

3. Автоматты сөздікке енгізілген барлық тілдердің аударма баламаларына ортақ болып келетін сөздік мақала құрастырылады.

Сөздіктің "керіқайтқыштық" қасиеті деген ұғымды тұтынушының қалауына қарай, екі түрлі тілдің қайсысын болса да автоматты жүйеге енетін немесе одан шығатын тіл деп санап, орындарын еркін ауыстыруға болатын мүмкіндігі деп түсіну қажет.

Көптілдік автоматты аударма жүйесінде керіқайтқыштық мәселесі екі түрлі деңгейде қарастырылады – тілдер деңгейінде және жеке лингвистикалық бірлік деңгейінде. Ондай жүйедегі керіқайтқыштықтың тілдік деңгейдегі түрінде сөздікке енген тілдердің қайсысын болса да *сұраныс тілі* ретінде және *жауап тілі ретінде* қарастыру мүмкіндігі бар. Ал лингвистикалық бірлік деңгейіндегі керіқайтқыштықты (лексикалық керіқайтқыштықты) бір ғана лексикалық бірлікті сұраныс және жауап ретінде қолдану мүмкіндігі деп ұғыну қажет.

Барлық ірі автоматты аударма жүйелері жарым-жартылай немесе толық түрдегі керіқайтқыштық қасиетке ие болып келеді. Ал автоматты аударма жүйесінің керіқайтқыштық қасиетке ие болмауы, оның аса үлкен кемшілігі деп саналады. Сонымен бірге олардың аударма терминологиялық сөздіктермен де көптеген ортақ қасиеттерге ие екендігі белгілі.

Автоматты аударма жүйесінің керіқайтқыштық қасиеті оны дәстүрлі түрдегі аударма лексикографиясындағы кейбір автоматтандыру жұмыстарында қолдануға мүмкіндік тудырады. Аударма сөздігін түзудегі уақытты үнемдеу сипаты ондағы тілдерді, сөздік мақалалардың жүйеге енетін және одан шығатын тілдері ретінде, автоматты түрде орындарын ауыстырып тұруына қатысты болады.

Барлық ірі автоматты аударма жүйелерінің негізгі ерекшелігі олардың көптілділігі. Мысалы, 20 жылдай бойы тәжірибеге енген LEXIS атты автоматты сөздік қазіргі кезде мынадай лексикалық топтарды өз құрамында сақтайды: неміс, ағылшын, француз, орыс, итальян, дат және португал тілдері. Сол сияқты, EURO DICAUTOM автоматты аударма жүйесі алты түрлі тілді қамтиды: ағылшын, дат, итальян, неміс, нидерланд және француз тілдері. Ал қазіргі кезде TFAM автоматты аударма жүйесі тоғыз түрлі тілді қамтып, олардың 2 млн көлемді тілдік бірліктерімен аударма жұмыстарын атқара алады. Кезіндегі Бүкілодақтық аударма орталығының (**ВЦП**) автоматты сөздігі және Дрезденгі техникалық университеттің терминоло-гиялық банкі өз құрамында төрт түрлі тілді қамтиды: ағылшын, неміс, француз және орыс тілдері.

Дәстүрлі аударма сөздіктеріне қарағанда автоматты сөздік динамикалық қасиетке ие, яғни олар сөздіктің сөзтізбесі мен сөздік мақаланың көлемділігіне қарамай, оларға қажетті түзетулерді тез және еш қиындықсыз енгізе алады.

Автоматты сөздікті құрастыру – ғалымдардың лексикалық топты таңдау жұмысымен тұрақты айналысуларының арқасында, сөздіктің халық алдында қызмет ету кезеңдерінің бәрінде де өз жалғасын тауып отырады. Қажетсіз деген сөздерді жүйеден шығарып тастап және оны жаңа лексикамен толықтырып, сөздік мақаланың кез келген тұстарын әрі түзетіп, әрі толықтырып отыру мүмкіндігі тұрақты түрде орын алады. Құрамының айнымалығы мен сөздіктің қызметі кезіндегі түрлі түзетулерді қабылдау мүмкіндігі – машиналық сөздіктердің тұрақты сипаттағы және дәстүрлі сөздіктерден айырым табатын бірден-бір сипаттамасы.

Соңғы жылдары аса кең орын ала бастаған сөздік – ЭЕМ арқылы құрастырылатын конкорданс-сөздіктер. Қазіргі кезде Шекспир, Ронсар, Пушкин және т.б. классиктердің шығармалары бойынша іргелі конкорданс-сөздіктермен компьютер арқылы танысуға мүмкіндік туындап отыр.

Сөзнұсқағыш сөздік дегеніміз әліпби ретімен реттеліп жазылған және сөздердің қандай мәтіндерде қолданылғанына сілтеме жасалатын сөз немесе сөзтұлғалар сөздігі. Конкорданс-сөздік аталған мәліметтерді өз бойында ұстайды және сонымен бірге ол сөздікте әрбір сөзге қатысты мәнмәтіндер де бірге қамтылып отырады. Түсіндірме сөздіктерге қарағанда конкорданс-сөздіктерде сөз мағынасын анықтауға қатысты түсініктеме беру және қажетті деген мысалдармен жабдықтау қарастырылмайды.

ЭЕМ көмегімен түзілген **конкорданс** лексикографты көптеген әрекеттерден босатады. Мысалы, оған жататындар: сөзтізбені таңдау мен құрастыру, мәнмәтін көртішкілер қорын жинақтау және т.б. Егер кез келген конкорданс-сөздігіне лексикограф күшімен қажетті деген толықтырулар жасалынатын болса, ондай сөздік түсіндірме сөздіктердің түр-түрін құрастыруға негіз бола алады.

Конкорданстарды құрастыру жұмыстарын талдай келе, мынадай тұжырым жасауға болады. Қазіргі кезде ең тиімдісі – арнайы және толық түрде автоматтанған сөздік-анықтағыш қызметін атқаратын "сұрақ-жауап" ұстанымында құрылған конкорданс типтес сөздікті жобалау және оны іске қосу. Осындай жұмыстың негізінде тұтынушының өзіне қажетті деген шығармалар мәтінін компьютер жадына бір-ақ рет енгізіп, одан кейін өз қалауынша автоматтанған жүйе ретіндегі конкорданска өз қалауынша "енуіне" мүмкіндігі бар. "Деректер банкісі" деп аталатын мұндай жүйелер мәтін ішіндегі ақпараттың тек қана дәлдігін, сенімділігін және дұрыстығын тексеріп қоймайды және сонымен бірге тиісті сөздің барлық қолданысы жайлы, әрбір сөзге (сөзтұлғаға) қатысты барлық мәнмәтін түрлері жайлы және шығармалар жиынтығы жайлы толық мәліметтер бере алады.

Мұндай автоматтанған сөздік-анықтағыш қызметі ежелгі жазба ескерткіштерге, жеке авторлардың классикалық әдеби шығармаларына, арнайы терминологиялық жүйе көрінісіне, әр тілде жазылған мәтіндерге барынша бейімделе алады.

Сөздік-анықтағыш қызметін атқаратын автоматтанған лингвистикалық деректер қорын құрудың өзектілігі еш күмән туғызбайды және мұндай мүмкіндіктің "ауылы" да аса алыс емес.

Сонымен, сөздік-анықтағыш жүйесін құру болсын немесе конкорданстар жүйесін құру болсын бұлардың бәрі де өзекті жалпы-филологиялық мәселе екені айқын. Мұндай жүйелерді ойдағыдай жүзеге асыру үшін тілшілердің, әдебиетшілердің, математика және қолданбалы лингвистика мамандарының тығыз түрде одақтасып, бірлесе жұмыс атқарулары қажет екені анық.

Терминологиялық жұмыстарды механикаландыру мен автоматтандыру қызметін ғылыми-техникалық әдебиеттер мен құжаттар мәтіндерін дәстүрлі тәсілмен өңдеу әрекеттерімен бірге, оларды жүзеге асыру бірнеше функционалды элементтерден тұратын терминологиялық банк құруды, яғни автоматты жүйе құруды қамтамасыз еткен болар еді. Бұл жүйелердің негізінде барынша аз мезгілде, оңай жолмен және ең төмен қаражатпен терминологиялық деректерді жинақтау, сақтау және тарату жұмыстарын іске асыруға болады.

Соңғы кезде Ресей мен бірқатар басқа да шетелдерде (Франция, Швеция, Бельгия, Канада) әртүрлі білім салалары бойынша деректердің құрама **терминологиялық деректер банкін** құруға байланысты жұмыстар кең түрде өрістей бастады. Келесі кезекте тұрған жұмыс түрі – банктар арасындағы телебайланыс қызметін іске қосу әрекеті.

Мәселен, Ресейдегі автоматтанған ақпараттық-терминологиялық жүйенің қызмет атқаруының негізінде – ғылым мен техниканың әртүрлі аясындағы терминдерді қамтитын "терминдердің анықтағыш банкін" құрастыру жүргізілуде. Бұл банкке қажетті ақпараттық дереккөзінің рөлін ғылыми-техникалық терминдер мен анықтамаларға арналған мемлекеттік және салалық стандарттар, терминологиялық халықаралық стандарттар, ұсынылған терминдердің жинақтары атқарады.

Негізінде, терминологиялық деректер банкі, бір жағынан, ұғымдар мен терминдер туралы және олардың қалай таңбалатыны жайлы мәліметтер беретін *құжаттық-фактографиялық ақпараттық жүйе* түрінде, ал екіншіден, арнайы мәтіндерге арналған *автоматтанған ақпараттық сөздік-анықтағыш* түрінде ұйымдастырылуы қажет.

Терминологиялық деректер банкін екі түрлі аспектіде қарастыруға болады. Ең алдымен, бұл автоматтанған ақпараттық жүйенің белгілі бір түрі деуге болады. Бірінші аспект бойынша терминологиялық деректер банкі ғылыми-техникалық лексикография тұрғысынан қарастырылатыны белгілі. Сондықтан мұндай көзқарасты ұстаным ететін болсақ, дәстүрлі терминологиялық сөздікті түзудегі мәселелермен айналысуға тура келеді. Яғни пәндік білім аясын таңдау мен оларды шектеу, ғылым тарихына қатынасы, тілдерді, материалдарды, сөзтізбені қалыптастыру әдісін және т.б. таңдап алу мәселелерімен айналысуды қажет етеді.

Автоматтандыру әдісі тек ғылыми-техникалық лексикографиядағы, әсіресе, терминологиялық сөздіктерді құрастыру ісіндегі жинақталған материалдарды өңдеу мен сұрыптау үшін ғана қажет емес, сонымен бірге сөздік мақаланы құрастыруда да аса қажет. Семантикалық жағынан қарастырғанда терминдер айтарлықтай анық топтарды құрайды деуге болады. Мысалы, оған жататындар: құрал-аспаптардың, станоктардың, тетік пен құрылғылардың және процестердің атаулары. Бұл аталғандардың өздері көптеген нақты түрлерге бөлінеді. Ал мұндай мәліметтер белгілі термин семантикасын сипаттауға қажетті өлшемдердің нақты моделін алдын ала ұсынуға мүмкіндік жасайды. Мысалы, көптеген аспап-құралдарды сипаттау үшін олардың мынадай сипаттары ескерілуі қажет: атқаратын қызметі, тұратын орны, бөлшектер құрамы, әрекет ету принципі мен зонасы. Ал биологиялық түрді сипаттауға қажеттілер – тегі, жынысы, морфологиялық және биологиялық ерекшеліктері, ареалы (тараған ауданы), шаруашылық қызметі. Аталған параметрлер жалпы және жекелік сипатта да айырым табады. Сонымен қатар әрбір білім саласының өзіне ғана тән қасиеттері, процестері, әрекет аясы және т.б. ажыратылады. Сипаттау параметрлерін олардың терминдерді анықтаудағы ұстайтын ұстанымы бойынша және ғылымның ерекшелігіне, сөздіктің арналуына қарай топтау жүргізуге де болады.

Бірақ терминологиялық деректер банкі, сонымен бірге – құжаттық-фактографиялық типтегі автоматтанған жүйе екені айқын. Міне, осының салдарынан сөздік түзуге, құжаттарды индекстеу әдістемесін анықтауға, компьютерлік бағдарламаларды жазуға, компьютер мен тұтынушы арасындағы диалогты іске қосуға қатысты мәселелермен айналысу керек болады. Аталған проблемалық мәселелердің ойдағыдай шешім табуына байланысты өндірістік және өндірісаралық терминологиялық деректер банкі – барынша терең, семантикалық жағынан алғанда күшті құжаттық-фактографиялық сөздік-анықтағыш түріндегі диалогтық жүйе.

Болашақта түсіндірме және аударма сөздіктерді, анықтағыштарды, рубрикаторлар мен классификаторларды құрастыру жұмыстарының және терминологияны қалыпқа келтіру, терминологиялық комиссия жұмысына қызмет ету, ғылыми-техникалық аударма жұмыстарының бірден-бір негізгі бөлігі –

құжаттық-фактографиялық терминологиялық деректер банкінің құрама түрі болатыны сөзсіз.

Қазіргі кездегі есептеу техникасының, қолданбалы лингвистика мен ақпараттану салаларының табыстары мен жетістіктері мұндай автоматтанған жүйелердің үнемді және тиімді болу мүмкіндіктеріне нұсқайды.

Негізінде терминологиялық деректер банкісінің жадына сөз-терминдер жайлы, соның ішінде сөз таптарына, сөз тарихына, мәтіндегі қолданысына сілтеме жасалғандарын қосып есептегенде, көптеген басқа да мәліметтерді енгізуге болады. Терминологиялық деректердің банкілер жүйесі терминдік сөзге қатысты мұндай әр деңгейдегі мәліметтерді анықтай отыра, термин табиғатында бірге қайнасып жатқан ұғымдық мәліметті де қамтиды.

Беляева Л.Н., Герд А.С., Убин И.И. Автоматизация в лексикографии// Прикладное языкознание. Учебник / Отв. ред. А.С. Герд. СПб.: Изд-во С.Петерб. университета, 1996. С. 318-333.

Лингвистическая концепция терминологического банка данных машинного фонда русского языка // Под ред. А. С. Герда. М., 1989.

Марчук Ю.Н. Проблемы машинного перевода. М., 1983. 232 с.

Марчук Ю.Н. Машинный перевод как информационная и технологическая реальность. М., 1988.

Перстнев И.П., Волкова И.Н. Банки терминологических данных. М., 1987.

Попов Э.В. Общение с ЭВМ на естественном языке.М.,1982. 360 с.

Ritchie G.D., Hanna F.K. Semantic Networks – a General Definition and a Survey // Information Technology. 1983. Vol. 2. № 4. P. 187-231.

Vauquios B. Aspects of Mechanical Translation in 1979: Conference for Japan IBM Scientific Program. Grenoble, 1979. 50 p.

Vauquios B. La traduction automatique a Grenoble // Documents de linguistique quantitative. № 24. Paris, 1975. 184 p.

Wang W.S. – L., *Chan S.W., Ts'ou B.K.* Chinese Linguistics and Computer // Linguistics. 1973. 118. P. 89-117.

РЕФЕРАТ ҚҰРАСТЫРУДЫҢ ӘДІС-ТӘСІЛДЕРІ

Ақпарат көлемін ықшамдауға қатысты аннотациялау мен рефераттау ұғымдары. Ғылыми және ғылыми-техникалық ақпарат көлемінің үнемі ұлғаю сипатында болуына байланысты, оны тұтынушының өзіне қажетті деген мәліметтерді түпнегіз құжаттар жиынтығынан дәстүрлі іздеу тәсілімен тез арада тауып алу әрекеті қиындыққа соқтыруда. Сондықтан түпнегіз құжат бойынша екінші нұсқадағы құжаттар жиынтығын құрастыруды қажет етеді. Әдетте, оларға библиографиялық сипаттаулар, аннотациялар, рефераттар және ғылыми аудармалар материалдарын жатқызуға болады. Екінші реттегі құжат – түпнегіз құжаттағы ақпарат мазмұнын ықшамдау немесе "қысудың" (компрессиялаудың) нәтижесі болып табылады. Мұндай қысу мен ықшамдау библиографиялық сипаттаулар үшін өте-мөте тиімді, себебі мәтін орнында тек олардың тақырыптарының аттары ғана қалып отырады. Ал ғылыми аудармалардың көлемін "қысудан" ұтатынымыз аз. Себебі, аударма құжаттағы ақпараттың көлемі мен оның мазмұны түпнұсқа құжатпен барабар болуы қажет. Соңғы кезде кейбір аудармалардағы айтылымдарды қысқарту арқылы көлемді ықшамдау әдісі тәжірибеге енуде. Бірақ мұндай әдіспен де көлемді ықшамдау тұрғысынан ұтатынымыз аз. Ақпарат мазмұнынан аннотация мен реферат алу кезіндегі мәтін көлемінің "қысылу" нәтижесі жоғарыда аталған библиография мен ғылыми аударма құжаттарымен салыстырғанда көлемдік жағынан олардың орта шенінде деуге болады.

Аннотация мен реферат бір-бірінен ақпарат мазмұнынан айырым табады. Аннотацияның қызметі – толығымен нұсқағыштық. Себебі, аннотацияда түпнегіз құжаттағы ақпарат мазмұнын толық ашпаса да, сол мазмұнға нұсқау тәсілі орын алады. Ал реферат қызметіне келетін болсақ, ол ақпараттың мазмұнына нұсқаумен бірге, қысқаша ақпарат беру рөлін де атқарады. Яғни түпнегіз құжаттың мазмұны рефератта (азды-көпті болса да) толық дәрежеде болуға әрекет етеді. Ағылшын тілді елдерде аннотация мен реферат (indicative abstract) құрастыруда ақпараттың көлемдік өлшемі алынады. Аннотацияның таза нұсқағыштық қызметі аз көлемді қажет етсе, ал реферат оған қарағанда (өз рөлін толық орындау үшін) әлдеқайда көп көлемді қажет етеді. Әсіресе мұндай жағдай гуманитарлық ғылымдарға қатысты рефераттарға тән. Алайда рефераттардың көлемін дәл анықтау мүмкін бола бермейді, себебі ол көптеген факторларға байланысты болып келеді. Мысалы, оған жататындар түпнегіз құжаттың сипаты (мақала, монография және т.б.), түсініктілігі (түсінуге қиын құжаттар толығырақ рефератталады), түпнұсқа құжаттың тілі (сирек пайдаланылатын тілдердегі құжаттар егжей-тегжейлі рефератталады), ғылыми типі немесе пәндік ая түрі (жаратылыстану ғылымы немесе қоғамдық ғылымдар аясы) және т.б.

Интеллектуалды рефераттау. Түпнұсқа құжат мазмұнын интеллектуалды рефераттау ұғымына адам баласының ой қызметінің әрекеті арқылы, әдетте, маман адам (референт) арқылы орындалатын рефераттау әрекетін атайды. Қазіргі кезде рефераттық журналдардың басым көпшілігінде осындай әдіс орын алады.

"Интеллектуалды" терминін бейнелейтін ұғымның негізінде рефераттау жұмысын жүзеге асыратын субъектінің адам екені мәлім. Сондықтан интеллектуалды рефераттау әдісіне ЭЕМ көмегімен атқарылатын автоматты рефераттау әдісі қарсы мағынада айтылып жүр және соңғы әдістің, яғни автоматты рефераттау әдісінің рөлі мен сұранысы барынша артуда.

Рефераттау мәселесін жеке қарастырғанда автордың өзі жазған рефераты (summary) мен сол саланың маман референтінің рефератын (abstract) ажырата білу аса маңызды. Бірінші рефератты түпнегіз құжат бойынша автордың өзі құрастырады және онда құжаттың қалам иесінің түсінігіндегі ең маңызды деген аспектілері бейнеленеді. Әдетте, авторлық реферат түпнұсқа құжатқа (мысалы, мақалаға немесе монографияға) қысқаша шолу ретінде көрініс табады. Бірақ мұндай реферат әр уақытта жасала бермейді және тіпті де барлық авторларға бірдей тән емес.

Авторлық рефераттың қызметі – мәтіннің мазмұны жайлы оқырманды ақпараттау болып табылады. Мұндай ақпараттау нәтижесі аннотация арқылы берілетін нұсқағыш ақпараттан аз айырым табады. Авторлық реферат ешбір іздестіру қызметін атқармайды. Ал әдеттегі рефераттық журналдардағы рефераттар іздестіру қызметін де атқарады. Себебі "қолмен іздеу" қызметі кезінде құжаттың релеванттық критерийі (мәндік өлшемі) рефераттың мазмұнына қатысты анықталады. Кейбір өндірісі дамыған капиталистік мемлекеттердің рефераттық журналдарында авторлық рефераттар көбінде ашық түрде жарнамалық сипатқа ие. Сондықтан авторлық рефераттар бойынша құжаттың негізгі мазмұны жайлы пікір айту қиынға түседі.

Авторлық рефераттар көлемі жағынан, әдетте, референттердің құрастырған рефераттары көлемінен айтарлықтай аз болады. Көпшілік жағдайларда мұндай рефераттарда зерттеу жұмыстарының нақты нәтижелері келтіріле бермейді.

Рефераттық журналдар (РЖ). Қайсыбір пәндік аяны немесе келелі мәселені бейнелейтін библиографиялық сипаттаулар олардың тақырыбы бойынша немесе автордың аты-жөні бойынша реттелуі мүмкін. Әдетте, мұндай әрекет *әдебиеттердің библиографиялық көрсеткіші* деп аталатын кітап түрінде дайындалады. Библиографиялық сипаттаулардың бәрінде бірдей (немесе тек кейбіреулерінде) аннотация берілуі мүмкін. Осының нәтижесінде әдебиеттердің аннотацияланған библиографиялық көрсеткіші құрастырылады.

Әрбір ғылым саласына қатысты библиографиялық сипаттаулар каталогтық көртішкілер түрінде де көрініс табуына жағдай жасалады. Каталогтық жәшіктерге бұл көртішкілер автордың аты-жөні бойынша немесе білім салалары арқылы әліпби тәртібіменен реттеліп қойылса, олар *авторлық* немесе *жүйелік каталогтар* деп аталады. Бұларды әдебиеттердің библиографиялық көрсеткішімен бірдей дәрежедегі қолмен орындалған (автоматтанбаған) *ақпарат-іздестіру жүйелері* деп те қарастыруға болады.

Рефераттық журналдардағы рефераттар да осы айтылғандарға ұқсас түрде жүйеленеді. Бұлайша жүйелеудің негізінде – білім универсумын (әлемдік білімді) қалаған дәлдік дәрежесін сақтай отыра, салалар мен пәндік аяларға бөлу принципі жатыр. Қажетті деген мәліметті тауып алу әрекетін барынша жеңілдететіндіктен мұндай рефективтік журналдар да қолмен жүзеге асатын (автоматтанбаған) *қарапайым ақпарат-іздестіру жүйесінің* қатарына жатады.

Қазіргі кезде, әлемдік деңгейде қарастырғанда, тақырыбы мен құрылымы жағынан өзгешеленетін көптеген рефективтік журналдар қолданыс тауып жүр. Мысалы, Кеңес Одағы кезіндегі рефективтік журналдардың негізгілері ретінде РЖ ВИНИТИ және РЖ ИНИОН түрлерін атауға болар еді Әрине, одан басқа да рефективтік журналдарды, мысалы, РЖ ВНИИМИ, РЖ ВИНТИСХ және т.б. осылардың қатарына қосуға болады.

РЖ ВИНИТИ жаратылыстану мен дәл ғылымдар құжаттарын, РЖ ИНИОН гуманитарлық ғылымдар салаларын бейнелейді. Рефективтік журналдардың деректері сыртқы пішін жағынан – форматы бойынша, ал іштей – рефераттардың құрылымы мен көлемі жағынан айырым табады. РЖ ВИНИТИ рефераттары РЖ

ИНИОН-ға қарағанда көлем жағынан барынша шағын келеді және канонды (тұрақты) қалыпқа ие болып келеді. Басқаша айтқанда, рефераттың бастапқы жағында түпнұсқа құжатында сипатталған мәселелерге нұсқау жасалады да, ал қорытынды бөлігінде алынған нәтижелер аталады. РЖ ИНИОН көлемі РЖ ВИНИТИ рефератының көлеміне қарағанда бірнеше есе көп (10 есе және одан да көп) және құрылымы жағынан шағын мақаладан ешбір айырмашылығы болмайды. Себебі, қоғамдық ғылымдар мәселесін бірнеше тармақтарға қисындастыру немесе шағын түрде қорытындылау мүмкін бола бермейді, ал ғылыми тұрғыда дәлелділік сипатының орын алуы аса маңызды деп саналады.

Рефераттық журналдарда кез келген құжаттың рефераты жеке-дара түрде берілмейді. Ол автордың аты-жөнін, бас тақырыпты және шығу деректерін қоса есептегендегі библиографиялық сипаттауды да бірге қамтиды. Мұндай жағдай библиографиялық сипаттау мәтіні мен реферат мәтінін әртүрлі тұрғыда реттеуге мүмкіндік тудырады. Рефераттардың ең көп тараған үш түрін атауға болады:

1) мынадай реттегі (ағылш. conventional abstract): бас тақырып – библиографиялық сипаттау – реферат мәтіні түріндегі әдеттегі реферат;

2) бірінші орында тақырыпқа қатысты сөйлем орналасқан реферат. Ақпараттың реті мынандай: тақырыптық сөйлем – реферат мәтіні – библиографиялық сипаттау. Мұндағы "тақырыптық сөйлем" дегенді, ақпараттық тұрғыдан қарастырғанда, ең маңызды деп саналатын реферат мәтінінен алынған сөйлем деп ұғыну керек;

3) бірінші орында "бас тақырып" орналасқан реферат. Оның реттілігі мынадай: бас тақырып – тақырыптық сөйлем – реферат мәтіні – библиографиялық сипаттау (бас тақырып қайталанбайды).

Әртүрлі рефективтік журналдарда осы аталған рефераттардың қайсыбір түрі қолданылады. Мысалы, "Chemical Abstract" рефективтік журналда рефераттың 2-типі, ал ВИНИТИ атты рефективтік журналда – рефераттың 1-типі қолданылады.

Рефераттау жайындағы әдебиеттерде интеллектуалды рефераттың ең ұтымды (оптималды) нұсқасын таңдауға қатысты мәселе бірнеше рет талқыланды. Оны реферат ішінде түпнегіз құжат мазмұнының қай құрамдас бөліктерінің қандай реттілікте көрініс табуының қажеттігі жайлы мәселе деуге болады. Бірақ көптеген тәжірибелердің нәтижелеріне сүйенсек, алдын ала берілген қайсыбір түпнегіз құжат рефератының ұтымды нұсқасы болмайды. Шындығында, түпнұсқа құжаттың бір ғана мәтіні бойынша әртүрлі рефераттар құрастыруға болады. Және мамандар тұрғысынан олардың бәріне де қанағаттанарлықтай деген баға берілуі мүмкін. Сондықтан аталған мәселеге соңғы кезде басқаша түрде түсініктеме беріліп жүр: рефераттың ең ұтымды нұсқасын іздестіруден гөрі, құжаттың бір ғана мәтіні бойынша белгілі бағыттағы мамандар қажетін қанағаттандыра алатын бірнеше реферат құрастырылғаны жөн. Құжат мәтінінен реферат алу мәселесінде осындай көзқарасты ұстаным ету автоматты рефераттау жүйелерін құрастыруда көптеп қолданыс таба бастады.

Автоматты рефераттау. Автоматты рефераттау жүйесін құрастырумен айналысатын көпшілік ғалымдар түпнегіз құжаттың мәтінінде ақпараттау фрагменттерінің қайсыбір түрі орын алады деген тезиске сүйенеді. Сондықтан автоматты рефераттаудың негізін құрайтын ең ақпаратты деген фрагменттерді айқындау зерттеу жұмысының басты міндеттерінің бірі болып табылады.

Қазіргі кездегі түсінікке сәйкес, ақпараттың көпшілігі мәтіннің мынадай тұстарында шоғырланады:

1) мәтіннің бір нүктесінде;

2) мәтіннің бірнеше фокалды немесе тірек нүктелерінде;

3) мәтіннің кез келген жерінде кездесуі мүмкін, сондықтан оның қатаң түрде анықталған орны жоқ.

Ең көп тараған көзқарас бойынша, өзінде барынша көп мөлшердегі ақпаратты сақтайтын мәтін орны – тірек нүктелер. Ондай нүктелер, әдетте, мәтіннің басында немесе соңында орналасады. Дәлірек айтсақ, ондай орындарға: тақырып, кіріспе, тұжырымдар, қорытынды, ұсыныстар жатады. Ал аталғандардың ішінен ең негізгісі деп тұжырымдар мен қорытындыны атауға болады. Бірақ бұдан да басқа пікір кездеседі. Мысалы, ол бойынша ең көп ақпарат бөліктері мәтіннің барлық жерінде азды-көпті болса да бірдей дәрежеде үлестіріледі.

Қандай көзқарасты ұстаным етсек те мәтіннің ақпараттық бөліктерін айқындау үшін барынша мұқият зерттеуден өткен әдіс-тәсілдер жиынтығы болуы керек. Қазіргі кезде мәтін мен оның бөліктерінің әртүрлі белгілеріне сүйенетін көптеген әдіс-тәсілдер қолданылып жүр. Бұлардың бәрі бірдей мәтінді үзінді бөліктерге бөлшектеуді ұстаным етеді: ең алдымен мәтін кесінділерге бөліктеледі, сосын олардың ішінен ең ақпаратты деген сөйлемдер, сөз тіркестері және сөздер іздестіріледі. Осындай әдіс-тәсілдердің кең қолданыс тауып жүрген түрлеріне жеке тоқталайық.

Статистикалық-дистрибутивті әдістер. Статистикалық-дистрибутивті әдістерді басқаша түрде – тілдік бірліктердің (морфема мен сөздердің) мәтін ішіндегі үлестірілу заңдылығын (тізбектеліп келу заңдылығын) статистикалық зерттеу әдісі деп те атайды. Аталған әдіс мына жайтқа негізделеді. Қарастырылып отырған тілдік бірліктің мүмкін болатын барлық қоршаулар жиынтығы мен олардың сандық деректерін (немесе статистикасын) дистрибуция (тіркесімдік) тұтастығы деп қарастыру. Ал тәжірибе арқылы алынатын статистикалық өлшем статистикалық-дистрибутивті әрекеттердің объективтілігі мен алгоритмделуін, яғни бұл әдіс-тәсілдердің автоматтану мүмкіндігін қамтамасыз етеді.

Статистикалық-дистрибутивтік әдіс-тәсілдер дешифровка жасауда (құпия жазуды тануда), статистикалық-комбинаторлық модельдеуде, автоматты индекстеу мен рефераттауда және семантиканы зерттеуде кең қолданыс тауып отыр.

Автоматты жолмен рефераттау мақсатында қолданылатын жалпы статистикалық-дистрибутивтік әдіс-тәсілдер ішінен бірнеше жеке әдістерді бөліп қарастыруға болады. Ең алдымен ондай жеке әдіске ғалым Х.П.Луннің ұсынған әдісін жатқызуға болады. Ол бойынша қарастырып отырған құжат үшін ең ақпаратты деген сөйлемдер қатарына бір-біріне барынша тығыз орналасқан мәнді сөздер жиынтығын қамтитын сөйлемдер жатады. Бұл жағдайда сөйлемдердің мәтіндегі орналасу орны ескерілмейді, сондықтан ақпаратты сөйлемдер түпнегіз құжаттың кез келген жерінен алынуы мүмкін. Сөйлем ішіндегі сөздердің шоғырлануы (топтануы) деп, қарастырып отырған мәтін үшін шеткі сөздері мәнді болып келетін кез келген сөзтұлға тізбегін атайды. Х.П.Лун әдісі бойынша, сөздер арасындағы мағыналық қатынас ескерілмесе де, ол толық автоматтандыру мүмкіндігімен ұтымды деп саналады.

Лун әдісін оңтайландыру мақсатында, оның бірнеше түрленімдері жасалған. Мысалы, В.А.Освальдтың әдісі бойынша, сөйлемнің ақпараттылық деңгейінде тек сөздердің мәнділік дәрежесіне қарай шоғырлану мүмкіндігі ескеріледі. Басқаша айтқанда, бұл әдісте тірек сөздердің топталуы ғана ескеріліп қоймайды, сонымен бірге осындай шоғырланудың сандық сипаты да есепке алынуы керек болады. Ал Л.Эрлдің әдісі бойынша, сөйлемдердегі сөздердің шоғырлану саны үшке тең немесе үштен көп болса ғана ондай сөйлемді ақпаратты сөйлем қатарына жатқызуға болады. ACSI-Matic әдісінде шоғырланған тірек сөздер жиілігі құжат ішіндегі сөздердің орта жиілігінен үлкен болатын шамасы ғана ескеріледі. Л.Дойл мен М.Квиллиан мынадай әдістемені ұсынды: сөйлемдерде терминдердің немесе тірек сөздердің

бірге кездесу мүмкіндігін есепке алу мен арнайы матрица арқылы өзара байланыста болатын элементтердің ұқсастық коэффициентін есептеу.

Мұндай "бірге кездесу" принципін ұстаным ету басқа да ғалымдардың (мысалы, А.Я.Шайкевичтің) зерттеулерінде кездеседі. Бірақ ол ғалымдар бұл әдіс-тәсілді рефераттауды автоматтандыру мақсатында емес, тек семантиканы зерттеу әдісі ретінде қолданғаны белгілі.

Статистикалық-дистрибутивтік әдіс мәтін ішінен ең ақпаратты деп саналатын сөйлемдерді қарапайым жолмен бөліп алып, оларды рефератқа енгізуге мүмкіндік тудырады. Мұндай сөйлемдерді іздестіру аясын шектеу мақсатында статистикалық-дистрибутивтік әдіс позициялық (орналасу тәртібін ескеретін) әдіспен бірге қолданылуы да мүмкін.

Статистикалық-дистрибутивті әдістің кемшілігі ретінде мәтін үзіндісінің ақпараттылығы мен олардың құрамындағы тірек сөздердің арасында тікелей және бірмәнді байланыстың болмауын айтуға болады. Әрине, мұндай байланысты анықтау ісі барынша күрделі мәселенің қатарына жатады. Сондықтан бұл әдіс, бір қарағанда, аса қарапайым болып көрінгенімен, зерттеушіні қалаған нәтижесіне жеткізе бермейді. Сонымен бірге тек қана статистика-дистрибутті әдістемеге негізделіп құрастырылған реферат әр кезде бірдей дәрежеде жеткілікті түрде қисынды сипатта көрініс таба бермейді. Аталған рефераттың интеллектуалды рефератқа қарағанда кемшілік жақтары жеткілікті. Сондықтан қажеттілікке қарай, статистикалық критерий басқа критерийлермен толықтырылып отыруы керек.

Мағыналық аспектіні қолдану әдісі. Статистикалық-дистрибутивті әдіс бөліп алынған мәтін үзіндісінің мазмұндық жағы туралы ешбір мәлімет бермейді. Сондықтан оның ақпараттық сипаты жайлы тек жобалап қана айтуға болады. Бірақ бөлікке бөлініп алынған мәтін үзінділерінің мән-мағыналық жағын арнайы жүргізілетін әрекеттер арқылы нақтылауға болады. Мұндай әрекет түріне мәтін үзіндісін мағыналық индикаторлар арқылы белгілеу барысын жатқызуға болады. Аталған индикаторларды көпшілік жағдайда мағыналық аспект деп атайды да, ал оған негізделетін рефераттау әдісін аспектілік әдіс дейді.

Әр автор әртүрлі аспектілер жиынын ұсынуы мүмкін, бірақ солардың ішінен жиі қолданылатындары: зерттеу пәні, мақсаты мен зерттеу әдісі, актуалдық, ұсынылған әдістің ерекшеленуі немесе оның шешімдері, артықшылығы, қолдану аясы, түйіндері, нәтижелері, ұсыныстар және т.б. Кейбір жағдайларда мұндай аспектілер семантикалық топтарға біріктіріледі.

Рефераттаудың аспектілік әдісінің құндылығы белгілі. Ол әдіс бойынша рефератқа белгілі мазмұндағы және ең ақпаратты деп қарастыруға болатын сөйлемдер мен мәтін үзінділерін енгізуге мүмкіндік туындайды. Көптеген зерттеушілердің бақылауынша, ең ақпаратты деп саналатын аспектілерге "шешімге жетудің ұсынылған вариантын" және "түйіндерін" жатқызуға болады. Маңызды аспектілер қатарына, көбінде, тақырыпта көрініс табатын мән-мағыналар жатады.

Аспектілік әдіс түпнегіз құжат мәтінінен мәнділік зоналарды бөліп алуға және олардың әрбіреуіне салмақтық коэффициент шамасын тіркеп жазуға мүмкіндік туғызады. Салмақтық коэффициенттің шамасына қарай, сол зона үзіндісін рефератқа енгізу не енгізбеу мәселесі шешіледі. Аспектілік әдісті статистикалық-дистрибутивті әдіспен бірге де қолдануға болады.

Аспектілік әдіс бойынша құрастырылған рефераттар бір-бірімен айтарлықтай байланыста болмаса да, тұтынушыға жеткілікті дәрежеде ақпарат беретін бірнеше мәтін үзінділерінен тұрады. Мәтін үзінділерінің байланыссыздық дәрежесін *мәтіндік байланыстар әдісін* қолдану арқылы төмендетуге болады.

Мәтіндік байланыстар әдістері. Мәтіндегі сөйлемдердің ақпараттық мәнділігі тек олардың бөліктерінің статистикалық сипаттары мен мазмұндық аспектілері арқылы ғана анықталып қоймайды, ол сонымен бірге басқа да сөйлемдердің байланыстық сипаты мен санына қатысты болады. Көптеген ғалымдардың (В.Е.Берзон, С.М.Приходько, Э.Ф.Скороходько, И.П.Севбо және т.б.) ұсынған автоматты рефераттау моделі осындай ұстанымда құрылған.

Енді ақпараты мол және мәнділігі жоғары деп саналатын сөйлемдерді таңдау әдісін қарастырайық. Бұл әдіс бойынша байланыстардың санын білдіретін қайсыбір шекаралық (пороговая) шама алдын ала берілуі қажет. Осы шаманы реттеу арқылы сөйлемдердің ақпараттық қанықтығын айқындауға мүмкіндік бар. Мұндай әрекет бойынша рефераттың көлемін өзгертуге болады. Әрине, мұндай жағдайда сөйлемдердің сипаты ескерілмеуі де мүмкін.

Бұл әдістің тағы бір жақсы нұсқасында мәтіндегі фразадан тыс бірліктердің (сверхфразовые единицы) *эксплицитті* және *имплицитті* түрлері өзара ажыратылып қарастырылады. Мұндай жағдайдан, әдетте, негізгі мазмұн фразадан тыс бірлікке *эксплицитті* болып келетін бірінші сөйлемнің үлесіне қатысы бар. Бұл сөйлем, көпшілік жағдайда, автосемантикалық болып келеді. Жоғарыда сөз болған *фразадан тыс бірлікке эксплицитті* келетін "бірінші сөйлемнің" ақпараттылығы сол бірліктің ұзындығының артуына қарай ұлғайып отырады.

Фразалық байланыс өлшемдерін ескеру ісі реферат нәтижесін барынша *байланыстық* және *толықтық* сипатта бейнелейді. Бірақ осындай жағдайдың өзінде де аталған реферат сапалық жағынан интеллектуалды рефераттан көп төмен тұрады.

Құрамдастырылған әдістер. Мәтіннің ең ақпаратты деген үзіндісін бөліп алуға байланысты қолданылатын әрбір жеке әдістің іштей өзіне ғана тән шектеулері болатыны белгілі. Рефераттау әрекеті кезінде тұтынушыны қанағаттандыратындай нәтижеге жету үшін әр әдістің тиімді жақтарын бөліп алып, оларға құрамдастыру әдісін қолданған жөн болады. Мәселен, зерттеушілердің барлығы бірдей мәтін ішіндегі сөздің немесе оның үзіндісінің кездесу жиілігін ескерудің маңыздылығына нұсқайды. Сол сияқты мәтін бойындағы сөйлем бірлігін немесе оның басқа бір үзіндісін, ал басқа бір жағдайда олардың мәтіндік байланыстылығын тірек етіп алу ұстанымының маңыздылығы да алдыңғы әдістен ешбір кем емес. Тағы бір ғалымдар тірек сөздердің аралық қашықтығын, тақырып құрамына енген тірек сөздерді және мәтін типін ескеру қажеттігін де баса айтады. Ал морфологиялық түрпат (форма) пен сөздің немесе сөз тізбектерінің синтаксистік қызметі жайлы факторлар өте сирек сөз болады.

Нәтижелік рефератқа енуге тиісті ақпараттық мәнді мәтін бірліктерін немесе үзінділерін бөліп алу әдістері белгілі реттілікпен қолданылуы қажет. Мәселен, лексиканың статистикасын анықтауға арналған әдістер ең ақпаратты деген сөйлемдерді немесе олардың жиынтығын мәтін ішінен бөліп алу үшін қажет. Сонымен бірге мұндай бөліп алу барысы шептік (позициялық) және аспектілік әдістер арқылы нақтыланып, түзетіліп отырулары қажет. Бірақ рефераттың мәтіндік қисындылығына фразааралық байланыс өлшемін қолдану арқылы ғана қол жеткізуге болады. Сондықтан рефераттау әдістерін қолдану тәртібі төменнен жоғарыға қарай бағытталған табиғи қозғалыс сызбасы бойынша орындалады. Ондай сызба мынадай реттілікке негізделеді: лексика – сөйлем – шебі (позициясы) немесе аспектілі келісілген сөйлем жиындары – фразааралық байланыстар – реферат. Бірақ бұл сызбаға қарама-қарсы бағытпен де, яғни жоғарыдан төменге қарай бағытталған сызба түрі де қолданылуы мүмкін. Ол бойынша алдымен түпнегіз құжаттың толық мәтініне бастапқы үзінділерге бөлшектеу тәсілі жүзеге асырылады да, сосын барып

үзінділер аясында орныққан орнына қарай немесе аспектілік жағынан сипатталған сөйлемдер бөлініп алынуы мүмкін. Ең соңында лексикалық деңгейге өту жолы қарастырылады. Аталған екі әдісті бір-біріне қарама-қарсы бағыттағы екі әдіс-тәсіл ретінде жарыстыра қолдануға да болады.

Интеллектуалды және автоматтау әрекеті бойынша құрастырылған рефераттар. Интеллектуалды рефератты құрастырудың мақсаты – тұтынушыны қажетті ақпаратпен қамтамасыз ету немесе оның іздестіру қажеттілігін қанағаттандыру. Басқаша айтқанда, рефераттарды пайдаланатын адам болғандықтан, яғни оның ақпараттық қажеттігін өтеу. Әрине, мұндай рефераттар басқа да мақсаттар үшін пайдаға асуы мүмкін, мысалы, құжаттарды қол күшімен немесе автоматты түрде индекстеуге, терминологиялық сөздіктер мен ақпарат-іздестіру тезаурустарын құрастыруға, мәтіндердің атрибуциясын (шығарма авторы, шыққан мезгілі мен орны) анықтау және т.б. мақсаттар үшін қолданыс табуы мүмкін. Бірақ бұл аталған мақсаттардың барлығы да интеллектуалды рефераттар үшін негізгі емес, жанама деректер.

Ал автоматты түрде рефераттаудың мақсаттары интеллектуалды рефераттау мақсатынан көпшілік жағдайда ерекшеленеді. Әрине, мұндай рефераттар да тұтынушының, яғни адамның мұқтаждығын қанағаттандыру үшін арналған болуы мүмкін. Бірақ машина арқылы жүзеге асатын рефераттардың мүмкіншілігі әзірше интеллектуалды рефераттардың мүмкіншіліктерімен салыстыруға келмейді.

Машиналық рефераттар көбінде индекстеу мақсатында қолданылады. Сонымен бірге мұндай мақсатқа байланыссыз немесе байланысы аз сөйлемдердің немесе олардың үзінділерінің тізбегі түрінде көрініс табатын рефераттар да сай келе беруі мүмкін. Аталған рефератқа қойылатын негізгі талап – құжатты индекстеуге қажетті тірек сөздер санының жеткілікті болуы. Мұндай рефераттар, әдетте, квазиреферат деген терминмен аталып жүр. Квазирефераттардың қолданылатын екінші бір саласы ретінде мәтіннің атрибутын атауға болады. Бұл бағыт бойынша мәтінге пәндік салаға сәйкес келетін индекстер тіркеліп жазылуы қажет. Бұл жағдайда рефераттың байланыстылық сипатының жоғары дәрежелі болуы қажет те емес.

Сонымен, квазирефераттар дегеніміз өзінше бір лексикалық тұрғыда өзекті саналатын рефераттар, ал интеллектуалды рефераттар – дискурс бойынша өзекті рефераттың түрі. Басқаша айтқанда, олар байланысты мәтін өлшемдерін қанағаттандыра алатын реферат түрі деуге болады.

Бұл аталған айырмашылықтар аса маңызды болғанымен, жоғарыда айтылғандардан машиналық рефераттың өлшемдерін интеллектуалды рефераттың параметрлеріне жуықтататын автоматтанған амалдар жүйесін құру мүмкін емес деген қорытынды шықпайды. Мұндай мақсатқа жету үшін автоматты түрде рефераттаудың әртүрлі әдіс-тәсілдер мен амалдар жиынтығын қолданумен бірге, олардың үйлесімділік жағын да ескеру қажет болады. Қорыта айтқанда, мұндай бағытты ұстаным ету – әрбір тұтынушының талабына сай келеді және қажетті деген прагматикалық бағытты қамтамасыз ете алады.

Бондаренко Г.В. Об использовании формальных методов анализа структуры целого текста для автоматического индексирования и реферирования // Научно-техническая информация. Сер. 2. 1975. № 5.

Гендина Н.И. Выявление аспектов содержания документов с помощью формальных тестовых признаков // Научно-техническая информация. Сер. 2. 1979. № 4.

Гиндин С. И. Методы автоматического фрагментирования текста, опирающиеся на характеристики внутреннего состава фрагмента // Семиотика и информатика. Вып. 9. М., 1977; Позиционные методы автоматического фрагментирования текста, их теоретико-

текстовые и психолингвистические предпосылки // Семиотика и информатика. Вып. 10. М., 1978.

Горькова В.И., Борохов Э.А. Реферат в системе научной коммуникации: Направления совершенствования лингвистических и структурных характеристик // Итоги науки и техники. Сер. Информатика. М., 1987.

Добрускина Э.М., Берзон В.Е. Синтаксические сверхфразовые связи и их инженерно-лингвистическое моделирование. Кишинев, 1986.

Карасев С.А. Построение системы автоматического реферирования, включающей анализ и синтез текста (АНАСИН) // Научно-техническая информация. Сер. 2. 1978. № 5; Система автоматического реферирования: результаты машинного эксперимента // Научно-техническая информация. Сер. 2. 1978. № 8.

Леонов В.П. О методах автоматического реферирования (США, 1958–1974) // Научно-техническая информация. Сер. 2. 1975. № 6.

Новиков А.И., Якушин Б.В. Алгоритм индексирования текстов взвешенными ключевыми словами по методу семантической фильтрации // Научно-техническая информация. Сер. 2. 1972. № 6. С. 15-20.

Пащенко Н.А., Кнорина Л.В., Молчанова Т.В. и др. Проблемы автоматизации индексирования и реферирования // Итоги науки и техники. Т.7. М., 1983.

Севбо И.П. Структура связного текста и автоматизация реферирования. М., 1969.

Скороходько Э.Ф. Семантические связи в лексике и текстах // Вопросы информационной теории практики. 1974. № 23.

Шайкевич А.Я. Дистрибутивно-статистический анализ в семантике // Проблемы и методы семантических исследований. М., 1976.

АҚПАРАТ ІЗДЕСТІРУ ЖҮЙЕСІ

Соңғы 20 жыл ішінде танымдық ой-сана мен өндіргіш күштердің дамуы нәтижесінде өмірге жаңа ғылым саласы – информатика ілімі келді.

Бір жағынан алғанда, информатика – ақпаратты ұйымдастыру мен өңдеудің жалпылама заңдылықтарын қарастыратын ғылым. Екіншіден, оны бүгінгі таңдағы қоғамдық өндірістің жаңа саласы деуге де болады. Информатика ілімінің өмірге келуі мен даму деңгейлері ең алдымен ақпаратты өңдеу құралдарының, оның ішінде есептеу техникасы мен байланыс құралдарының дамуына тікелей қатысы бар. Ғылыми-техникалық прогресті жаңаша түрде ақпаратпен жабдықтандыру мәселесі автоматтанған жүйелердің кең түрде қолданыс табуымен сипатталады. Қазіргі кезде деректерді электрондық әдіспен өңдеуге қатысты жүйелер әртүрлі мәселелерді шеше алады: есептеу, технологиялық тіркеу, статистикалық тіркеу, баспа жұмыстарының есеп-қисаптары және таза ақпараттық (ақпараттық сақтау мен іздестіру) мәселелер.

Ғылыми-техникалық прогрестен туындайтын ең бір маңызды салдар – ақпарат көлемінің бұрын-соңды болмаған дәрежеде артуы. Ғылыми-техникалық басылымдар мен жарияланымдардың сан жағынан өсуі қажетті деген мәліметтерді тауып алу мәселесін кенеттен қиындатып жіберді. Осыған байланысты ақпарат-іздестіру теориясы дамып, ақпарат іздестіру жүйелері мен құрылғыларды құрастыру жұмысында жүзеге асты. Ақпарат іздестіру жүйелерінің қол күші арқылы іске асатын механикаландырылған және автоматтандырылған түрлері болады. Ең алғашқы ақпарат іздестіру құрылғылары **сұратуға** (запрос) сай қажетті деген құжатты тауып алуды механикалық сәйкестендіру арқылы іске асыратын техникалық құрал түрінде жасалды. Оларда перфокарталар, суперпозициялық карталар, шеттері перфорацияланған перфокарталар және т.б. қолданыс тапты. Ақпарат іздестіру қызметінде қолданылатын есептеу техникасының дамуы *автоматтанған ақпараттық-іздестіру жүйесін* өмірге әкелді. Телекоммуникациялық құралдардың дамуы үлестірілген ақпараттық іздестіру жүйелерін, әлемдік ақпарат іздестіру жүйелерін және әлемдік ақпараттық инфрақұрылым тұтастығын жасауға өз әсерін тигізді. Біздің баяндауымызда тек формалды түрдегі сұратулар бойынша іске асатын, қажетті мәліметті бойында сақтаған мәтінді (құжат, оның бөліктері, фактографиялық жазба) іздестіруге арналған автоматтанған ақпараттық жүйелер тобы қарастырылады. Аталған жүйенің басқа жүйелерден негізгі айырмашылығы мынада. Автоматтанған ақпарат іздестіру жүйесі әрекетінің нәтижесінде алынған деректердің мағыналық мәнін ашу мен ол бойынша тиісті шешім қабылдау жақтары қарастырылмайды. Яғни ақпарат іздестіру жүйесі – мәтіндерді (құжаттарды) немесе мәліметтерді (фактыларды) сақтау мен іздестіруге арналған құралдар мен әдістердің жиынтығы болып табылады.

Ақпарат іздестірудің мәні және негізгі терминдер. Ақпарат іздестірудің мәні – *құжаттың мазмұны* мен *сұрату бейнесін* өзара теңдестіру. Ал екі түрлі мәтіннің мазмұнын салыстыру дегеніміз мәтіндердің мағыналық жақындық (ұқсастық) дәрежесін анықтау. Ақпарат іздестірудің негізгі ұғымы (құжат пен сұрату ұғымдарымен бірге) релеванттылық ұғымы (ағыл. relevant – *қолайлы, қатысы бар*) болып табылады. Әдетте, **релеванттылық** ұғымын бір мәтіннің екінші мәтінге сәйкестік дәрежесі – құжат пен сұратудың сәйкестігі деуге болады. Ал релеванттылық –

мағыналық релеванттылық пен сұратудың сәйкестігі ретінде де түсінік табады. Релеванттылық сонымен бірге мағыналық релеванттылық және формалды релеванттылық түрінде де екіге ажыратылады. *Мағыналық релеванттылық* – ақпаратты сұратуға байланысты табылған құжаттың ақиқаттық сәйкестігі. Ал *формалды релеванттылық* – құжаттың іздестіру бейнесінің іздестіру ұйғарымына сәйкестігі. Бұл жердегі *іздестіру бейнесі* мен *іздестіру ұйғарымының* түсінігі дегенді құжаттың негізі мазмұны мен сұратуды арнайы ақпаратты іздестіру тілінде *бейнелеу* деп ұғынған жөн. Жоғарыда аталған мағыналық және формалды релеванттылықпен бірге "пертиненттілік" (ағыл. pertinent – *қолайлы, іске қатысты*) түрін де ажыратады, яғни құжат мазмұнның шындығындағы ақпараттық мұқтаждыққа сай келуі. Ал ақпараттық мұқтаждық – белгілі индивидтің немесе ұйымның нақты мақсатқа жету жолындағы білімге мұқтаждығы. Ақпараттық мұқтаждық табиғи тілдегі мәтін түрінде көрініс табады және ол *ақпараттық сұрату* деп аталады. Ақпараттық сұрату ақпараттық мұқтаждықтың көрінісін әрқашан бірдей дәрежеде дәлме-дәл сипаттай алмайды. Себебі, ақпараттық мұқтаждықты тұжырымды етіп айтудың субъективтік сипатта болуы әртүрлі қиындықтарға әкеледі.

Релеванттылық ұғымына ақпаратты іздестірудің негізгі нәтижелік көрсеткіші, яғни толықтық пен дәлдік сипаттары негізделеді. *Іздестірудің толықтық өлшемі* деп ақпараттық жиымда (массив) қамтылған сұраныска сай келетін релеванттық құжаттар санының құжаттардың жалпы санына қатынасын атайды. Ал *іздестірудің дәлдік өлшемі* – берілген релеванттық құжаттар санының тауып алынған барлық құжаттар санына қатынасы.

Іздестірудің нысаны мен алға қойған мақсатына байланысты екі түрлі ақпарат іздестіру орын алады: құжаттық және фактографиялық. Осыған сәйкес ақпарат іздестіру жүйесінің де құжаттық және фактографиялық түрлері айырым табады. Құжаттық іздестіруде тұтынушыға ұсынылатын нысандар ретінде қажетті мәліметтерді өзінде сақтайтын құжаттың өзі немесе олардың библиографиялық баяндалуы (описание) жайлы мәлімет беріледі. Ал *фактографиялық іздестіру* түрінде ақпаратты сақтау мен ұсыну нысаны ретінде дұрыс мәліметтер (нысандар, процестер, құбылыстар және т.б. ғылыми, техникалық, экономикалық қасиеттер) көрініс табады. Құжаттық және фактографиялық жүйелер арасында айтарлықтай айырмашылық болмайды. Сондықтан аралас жүйелер ұғымын да атап кетуге болады. Мұндай жүйелерде құжаттық және фактографиялық ақпараттар бір ғана нысанаға қатысты болады.

Құжатты іздестіруде фактографиялық ақпарат қосымша мүмкіндік ретінде қолданылады. Фактографиялық элементтерге ие болатын құжаттық іздестірудің бір түрі ретінде *библиографиялық іздестіруді* атауға болады. Мұндай іздестіру тәсілі құжаттың тек библиографиялық элементтерінің баяндалуы бойынша ғана іске асады (шыққан жылы, басылым орны, басылым түрі, авторы және т.б.). Әдетте, құжаттық ақпарат-іздестіру жүйеде сақталатын мәліметтер рефератпен қоса есептелетін библиографиялық көріністе беріледі. Ондай жиымдарға жасалатын *сұрату бейнесі* көбінде фактографиялық сипатта болады.

Ақпарат іздестірудің негізінде құжаттағы ақпарат бейнесін сұратудағы ақпараттың бейнесімен теңестіру әдісі жатыр. Сондықтан олардың мазмұн көрінісін іздестіру жүйесіне "жіберу" мәселесін шешіп алу қажет болады.

Ол үшін құжат пен сұратудағы негізгі тақырыпты және олардың формалды сипаттамаларын баяндауға арналған мамандандырылған арнайы тіл – *ақпарат іздестіру тілі* құрастырылуы қажет. Қазіргі кездегі автоматтандырылған ақпарат іздестіру жүйелерінде қолданыс табатын барлық ақпарат іздестіру тілін мынадай үш типке бөліп қарастыруға болады:

1) ұғымдарды жүйелі түрде топтастыруға негізделген – ақпарат іздестіру тілінің топтастыру типі;

2) сөз немесе сөзтіркестер арқылы көрініс табатын лексикалық бірліктердің тізбесіне негізделген – ақпарат іздестіру тілінің дескрипторлық типі;

3) белгілі заттық аядағы нысандар "белгі-мағына" мазмұнындағы білім моделіне негізделген – ақпарат іздестіру тілінің нышанды-белгілік типі.

Құжат немесе сұрату мазмұндарына көрініс беру әрекетін **индекстеу** деп атайды. Ақпарат іздестіру тілі арқылы сипаттау (жазу) кезінде құжат мазмұны қорытынды түрінде және "қысылған" пішінде көрініс табады. Ақпарат іздестіру тілінде жазылған құжаттың негізгі мазмұны құжаттың іздестіру бейнесі деп аталады. Ақпараттық мұқтаждықтың табиғи тілде жазылған пішіні – *ақпараттық сұрату* деп, ал оның ақпарат іздестіру тіліндегі көрінісі – *сұратуды іздестіру бейнесі* немесе *іздестіру ұйғарымы* деп аталып жүр.

Сонымен, ақпаратты іздестіру дегеніміз іздестіру жиымындағы іздестіру ұйғарымына сәйкес келетін құжаттың іздестіру бейнесін анықтауға әкеп соғады. Сәйкестіктің дәрежесін айқындау арнайы формалданған ережелер арқылы іске асады. Олар "шығарылу" (выдача) критерийі (өлшемі) немесе мағыналық сәйкестіктің критерийі (өлшемі) деп аталады. Бұл жерде тек формалды ережелер мен формалды релеванттық жайлы ғана сөз болып отырғандықтан, оны мағыналық релеванттылықпен ешқашан шатастыруға болмайды.

Шығарылу критерийінің ең қарапайым түрі ретінде құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің толық сәйкестігі немесе, дәлірек айтқанда, сұратудың іздестіру бейнесінің толық түрде құжаттың іздестіру бейнесіне қатысты болуы саналады. Бірақ, әдетте, іс жүзінде құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің жарым-жартылай сәйкестік критерийі (өлшемі) деп те қолданыс табуда.

Шығарылу критерийіне қойылатын талап, көбіне, оның тікелей іздестіру ұйғарымына қатысты болуы, сондықтан бір ғана ақпаратты іздестіру тілінің екі түрлі мүмкіндігі жайлы немесе бұл жерде құжаттардың және сұрыптаудың ақпарат іздестіру тілдері туралы айтуға болады. Бұл жағдайда сұрату тілі іздестіру мен шығарылу категориясына қатысты қосымша түрдегі арнайы мүмкіндіктерді өз құрамына қосып алады.

Ақпаратты іздестіру жүйесі үнемі адам мен машина аралығындағы жүйе немесе "адам-машина" жүйесі болып саналады. Сондықтан абстракты ақпарат іздестіру жүйесі (логика-семантикалық құрал) мен толық ақпарат-іздестіру жүйесінің (ақпараттық-іздестіру қызметі) түрлері деген атпен айырым табады. Осыған қарай, ақпараттық қызметтің: "сұрату – жауап" пен таңдалып алынған ақпаратты тарату жолдары ажыратылады. Бұлардың біріншісі бойынша өткен кезді шола қамтитын жиым құжаттары негізінде бір реттік сұратулар алдын ала қарастырылады. Сондықтан мұндай іздестіру тәсілі, әдетте, *"ретро"* (ретроспективный – *өткенді шолатын*) деп аталып жүр. Ал екіншісі бойынша – таңдалып алынған ақпаратты тарату жолында – жаңадан түскен құжаттар жиымы негізінде тұрақты (белгілі мезгіл ішінде ғана күші бар) сұратулар алдын ала қарастырылады. Орыс тілінде сұратудың мұндай тәсілі – *"профиль"* деп аталып жүр.

Адам мен ЭЕМ арасындағы "тілдесуге" қатысты "пакетті" және "диалогты" тәртіп жолдары өзара өзгешеленеді. Біріншісі, яғни "пакетті" тілдесу таңдалып алынған ақпаратты тарату тәсіліне тән болса, екіншісі – диалогтық тәртіпке – "ретро" тәсіліне тән. Пакеттік тәсілде іздестіру жолы адамның қатынасынсыз, әдетте, барлық сұратулар бойынша бір мезгілде орындалады.

Диалогтік тәсілде іздестіру кезіндегі аралық нәтижелер ескеріліп, қажетті мәселенің шешімі табылғанға дейін тұтынушы (немесе арнайы дайындалған маман) ЭЕМ-мен диалогке түседі, яғни сұрату мен іздестіру стратегиясын нақтылай түсетін шешім қабылдайды. Іздестірудің қай жолы болса да, тұтынушыға ұсынылған нәтиже – құжат, немесе екінші дәрежелі (библиографиялық баяндау, мүмкін, референтпен бірге), немесе бірінші дәрежелі қажетті құжат болуы мүмкін.

Ақпаратты іздестіру жүйелерін олардың функционалдық белгілеріне қарай, мынадай құрама бөлшектерге (ішкі жүйелерге) бөледі: құжатты енгізу, сұратуларды енгізу және өңдеу, іздестіру, сөздіктерді түзу, статистиканы анықтау, іздестіру нәтижелерін өңдеу және қажетті деген құжатты ұсыну.

Аталған қызметтерді іске асыратын құралдар мен әдістер *ішкі жүйелерді жабдықтау* деп аталып жүр. Әдетте, ішкі жүйелерді жабдықтаудың мынадай негізгі түрлерін атауға болады: *ақпараттық жабдықтау; лингвистикалық жабдықтау; техникалық жабдықтау; программалық жабдықтау; технологиялық жабдықтау.*

Ақпараттық жабдықтау – құжаттар мен сұратулардың ақпараттық жиымы және оларды сипаттаудың, құрастыру мен топтастырудың құралдары мен тәсілдері.

Лингвистикалық жабдықтау – ақпарат іздестіру тілінен, индекстеу ережелері мен ұсыну критерийінен тұратын логика-семантикалық аппарат.

Техникалық жабдықтау – ақпаратты сақтау мен іздестіруді қамтамасыз ететін техникалық құралдар.

Программалық жабдықтау – ЭЕМ арқылы ақпарат іздестіру жүйесінің барлық функцияларын іске асыратын программалық мүмкіндіктер.

Технологиялық жабдықтау – автоматтанған және автоматтанбаған процестер мен ақпарат-іздестіру жүйесінде ақпаратты өңдеу әрекеттерінің баяндалуы (описание).

Ақпарат іздестіру жүйені лингвистикалық жабдықтау дегеніміз ақпаратты іздестірудің негізгі міндетін орындауға арналған логикалық және лингвистикалық құралдар мен әдістердің жиынтығы. Лингвистикалық жабдықтаудың құрамы, ең алдымен, тілдік құралдардан тұрады – арнайы тілдер (оның ішінде шектеулі табиғи тілді қоса есептегенде), осы тілдерге аударудың ережелері және оларды қолдану әдістері. Ақпарат іздестіру тілін жүзеге асыру үшін әртүрлі сөздіктер, қалыптар (формулярлар) және лингвистикалық алгоритмдер құрастырылады. Лингвистикалық жабдықтаудың ұғымына ақпарат-іздестіру жүйесінде қолданылатын шығарылу критерийі, математикалық логика мен жиындар теориясының құралдары да бірге енеді.

Лингвистикалық жабдықтауға негіз болатын – *ақпарат іздестіру тілі.* Адамдардың қызметі жағдайында танымдық және коммуникативтік функцияларды атқаратын көптеген басқа да таңбалық жүйелерге қарағанда, ақпарат іздестіру тілі алғашынан-ақ формалданған ақпаратты іздестіруге арналған. Осы тұрғыдан ақпарат іздестіру тілінің айрықша ақпаратты-іздестіру қызметін сөз етуге болады.

Ең жақсы деген ақпарат іздестіру тіліне мынадай негізгі талаптар қойылады:

1) жеткілікті түрдегі "семантикалық күштілік", яғни құжаттар мен сұратулардағы негізгі тақырыпты (тақырыптарды) толық және дәл сипаттау мүмкіндігі;

2) бір мағыналық, яғни ақпарат іздестіру тіліндегі әрбір жазбаның (запись) бір ғана мағынаға және әрбір ұғымның ақпарат іздестіру тілі арқылы біркелкі түсінікке ие болуы;

3) ашық жүйе сипатында болуы, яғни тілдің лексикалық құралдарын үнемі толықтырып, өзгертіп, түзетіп тұру мүмкіндігі.

Кез келген жазба тіл, ең аз дегенде, үш түрлі компоненттен тұрады: әліпби, сөздік және грамматика. Бұл жағдай ақпарат іздестіру тіліне қатысты да орын

алады. Әртүрлі ақпарат іздестіру тіліндегі лексикалық бірліктерді жазуға арналған графикалық белгілер жүйесі, яғни оның әліпбиі өз құрамында латын және орыс әліпбиін, араб цифрларын, тыныс белгілерін және арнайы символдарды сақтайды. Ақпарат іздестіру тілінің тілдік бірліктерінің жиынтығы *сөздік* деп аталады және тілдік бірлік ретінде ұғымдар мен ұғымдар топтары да бірге саналады.

Тілдік бірлік араларындағы қатынастарды анықтайтын жиынды *ақпарат іздестіру тілінің грамматикасы* деп атайды. Бұл қатынастар екі түрлі болып кездеседі: *синтагматикалық* (мәтіндік) және *парадигматикалық* (базалық). *Синтагматикалық қатынастар* дегеніміз нақты мәнмәтінде (яғни осы құжатта немесе сұратуда) белгілі бір жағдаятта пайда болатын тілдік бірліктердің араларындағы қатынастар.

Парадигматикалық қатынастардың туындауы үшін ақиқат әлемінің нысандары мен құбылыстары бейнелендіретін элементтердің орнына жүретін ұғымдар арасындағы логикалық байланыстың шарттылығы алынады. Табиғи тілдегі шағын ұғымдық түсінікте, грамматика дегеніміз қатынастардың бірінші типіне ғана қатысы бар мәселелер. Синтагматикалық қатынастарды белгілей алатын құралдары бар ақпарат іздестіру тілін "грамматикалы ақпарат іздестіру тілі", ал мұндай құралдары жоқ ақпарат іздестіру тілін "грамматикасыз ақпарат іздестіру тілі" деп атайды. Парадигматикалық қатынастар тілдің сөздік құрамына қатысты болады. Табиғи тілдердегі сөздіктерге қарағанда, ақпарат іздестіру тіліндегі сөздікте парадигматикалық қатынастар нақты түрде көрініс табады.

Координация ұғымына негізделген ақпарат іздестіру тілінің жалпы классификациясын сөз ететін болсақ, жоғарыда айтылғандай, ақпарат іздестіру тілінде тілдік бірлік ұғымдары, дәлірек айтсақ, ұғымдардың кластары орын алады. Ақпарат іздестіру тілі теориясында қарапайым ұғымдар кластарының қиылысулары арқылы құрастырылған күрделі ұғымдар кластарын *координация* деп атап жүр. Аталған теорияда ақпарат іздестіру тілінің құрылымы мен индекстеу ережелеріне қарай, *алдын ала координацияланған* (предкоординированный) және *соңынан координацияланған* (посткоординированный) деген ұғымдар бар. Оның біріншісінде, құжаттарды индекстеу, оларға алдын ала құрылған күрделі кластарды тіркеп жазуға негізделеді де, ал ол кластардың аттары іздестіру бейнесіне айналады. Егер іздестіру бейнесі қарапайым кластардан тұрып, ал координация амалы (күрделі кластарды қарапайым кластардың қиылысуы арқылы құру) тек нақты іздестіру кезіне ғана сай келсе, ондай ақпарат іздестіру тілін *соңынан координацияланған* деп атайды.

Аталған жағдайларды ескере отырып, ақпарат іздестіру тілінің жалпы классификациясын мынадай түрде беруге болады:

1. Алдын ала координацияланған ақпарат іздестіру тілі.
1.1. Иерархиялық классификациялар.
1.2. Әліпбилі-пәндік классификациялар.
1.3. Фасетті (біржақты) классификациялар.
2. Соңынан координацияланған ақпарат іздестіру тілі.
2.1. Дескрипторлық тілдер.
2.2. Семантикалық тілдер.
2.3. Синтагматикалық тілдер.

Алдын ала координацияланған ақпарат іздестіру тілінде барлық кластардың аттары да алдын ала беріледі. Әрбір атауға қатысты табиғи тілдегі оған сәйкес келетін ұғымның атауы беріледі.

Ең көп таралған иерархиялық классификация. Мысал ретінде әмбебап ондық классификацияны, ГАСНТИ-дың рубрикаторын, ойлап табудың халықаралық

классификациясын және т.б. атауға болады. Бұлайша классификация жасауға ұғымдарды бөлшектеу тәсілі негіз болып отыр. Классификацияда әрбір класс арнайы код арқылы белгіленіп, олар тұрақты және нақты орынға ие болады. Ақпаратты іздестіру тілі тұрғысынан қарастырғанда иерархиялық классификацияның негізгі кемшіліктері мыналар:

1) көпаспектілі индекстеуді қамтамасыз ете алмауы;

2) бөліктеу тереңдігінің жеткіліксіздігі;

3) ғылым мен техниканың дамуынан әрдайым артта қалып отыруы.

Иерархиялық классификациялау негізінен кітапханалар мен кітап қорларындағы алуан түрлі әдебиеттерді жүйеге келтіру жұмыстары үшін қолданылады.

Әліпбилі-пәндік және фасетті классификациялар нақты пәндік салада кездесетін негізгі категориялары мен пәндік кластардың алдын ала құрастырылған тізіміне негізделеді.

Әліпбилі-пәндік классификацияларда ұғымдар атаулары табиғи тілдегі сөздермен аталып, әліпбилік тәртіппен орналасады. Мұндай классификациялар шағын құжаттарды іздестіру жұмыстарына арналады. Бұлардағы негізгі тақырып сәйкес пәндік атпен белгіленіп, негізінен жүйелі каталогтерге нұсқағыш (көрсеткіш) құрастыру жұмысында қолданылады.

Фасетті классификация негізінде бір ғана жиындағы заттарды әртүрлі негізде топтау принципі жатыр. Дәлірек айтсақ, бір ғана жиын нысандары үшін бірнеше иерархиялық таралымдар құрылады. Фасетті классификациялау көпаспектілі индексация жүргізуге мүмкіндік тудырады. Негізгі категориялар – *классификациялау негіздемесі* – *фасет* деп аталады (ағылш. faset – *аспект*). Айталық, әртүрлі материалды белгілейтін терминдер жиыны "материалдар" фасетін құрайды, процестерді белгілейтін терминдер жиыны – "процестер" фасетін және т.б. құрайды. Фасеттің әрбір терминін *фокус* деп атайды және қолайлылық үшін оны қайсыбір код арқылы (цифрмен) белгілейді. Фасеттердің жиынын *классификация сызбасы* деп те атайды.

Мысалы, кинофильмдерді классификациялау үшін арнайы төрт фасетті бөліп алуға болады. Олар: жанр, метраж, түс және формат. Әрбір фасетте фильмді нақты аспект көзқарасында ғана сипаттайтын терминдер аталады. Сонда, кез келген фильмнің іздестіру бейнесі төрт фасеттен таңдалып алынған төрт термин-белгіден тұратын болады. Классификация сызбасында және оған сәйкес құжатты іздестіру бейнесінде фасеттердің орналасуы тұрақты белгіленеді. Мұндай орналасу жағы реттеліп, белгіленген тізбені фасеттік формула деп атайды. Сонымен бірге фасеттік классификациялау жұмысының үлкен қиындық туғызатындығынан, олар тек шағын пәндік сала үшін ғана жүргізіледі. Фасеттік классификацияны қолдану арқылы жүргізілетін индексациялау әдістемесі де аса күрделі. Сондықтан автоматтандырылған ақпарат-іздестіру жүйесінде аталғандай классификациялау іс жүзінде кең қолданыс таба алмауы мүмкін.

Сол сияқты семантикалық және синтагматикалық тілдердің де практикалық қолданысы жоқтың қасы. Бұлардың ішінен кең танымал болған Перри – Кенттің семантикалық коды, RX – кодының тілі, СИНТОЛ – тілі. Аталған тілдерге тән ерекшелік – ұғымдар мен айтылымдарды нақты сипаттайтын құралдың болуы. Ондай құралдардың орыс тіліндегі атаулары: "семантические множители, термы, реляторы, предикаты, сущности" және т.б.

Дескрипторлық тілдер. Автоматтандырылған ақпарат іздестіру жүйелердің даму сипаты, негізінен, *дескрипторлы типті ақпарат іздестіру тіліне* қатысты болады.

Дескрипторлық тілдерді құрастыруға координаттық индекстеу принципі негіз етіп алынады. Бұл ұстаным бойынша құжаттың негізгі тақырыбы – дескрипторлардың тізбесі, яғни қарапайым ұғымдардың аттарын білдіретін сөздер арқылы сипатталады. Мұндай сөздер қайсыбір ойша құрылған "n"-өлшемді пәндік-тақырыптық кеңістіктегі координаттардың орнына жүреді. Координаттық индекстеу принципі 50-жылдары механикаландырылған ақпарат іздестіру жүйесін құрастыру жағдайында пайда болған. Мұндай жүйелердің ең алғашқыларының бірі деп АҚШ-та М.Таубе құрастырған УНИТЕРМ жүйесін айтуға болады. Бұл жүйеде құжат пен сұрату мазмұндарын баяндау индекстері ретінде мәтін ішінен таңдалып алынған тірек сөздер (ключевые слова) қолданылады. Мұндай сөздердің (УНИТЕРМДЕР) барлығы да бірдей иерархиялық дәрежеге ие. УНИТЕРМДІК тілдердің сөздік құрамы алдын ала емес, тек индекстеу процесі кезінде тікелей құрастырылады. Шағын ұғымды анықтайтын координация немесе ұғымның логикалық көбейткіші екі реттен орын алатын тәрізді: құжатты индекстеуде – потенциалды (ойша) түрде және іздестіру процесінде – ақиқаттық түрде. Іздестіру процесі құжаттың іздестіру бейнесін іздестірудің алдын ала жазылып қойылған терминдерімен сәйкестендіру арқылы іске асады.

Мысалы, "кітапхана" және "автоматтандыру" ұғымдарының қиылысуы негізінде тар шеңбердегі жаңа ұғым – *"кітапханаларды автоматтандыру"* ұғымы туындайды.

Құжаттар мен сұратуларды тірек сөздер, яғни индекстеуге қажетті мәтіннен алынған терминдер арқылы индекстеу кезінде ақпарат іздестіру тілін лексикалық бақылау қажеттігі туындайды. Лексикалық бақылау – координаттық іздестіру кезіндегі сөздердің көпмағыналығы мен біркелкі қолданбауының теріс әсерін барынша азайту мақсатында қолданылатын шаралар жиынтығы. Лексикалық бақылаудың негізгі мақсаты – мәтіндегі берілу тәсіліне қарамай, құжат пен сұрыптаудың іздестіру бейнелері ұғымдарының бірмағыналы түсінігін қамтамасыз ету. Бұл жердегі лексикалық бақылау термині *сөздіктік бақылау* сөз тіркесімен синонимдес. Сондықтан ақпарат іздестіру тілінде лингвистикалық бірліктерді нормаға келтіру үшін барлық маңызды тірек сөздер мен дескрипторлардың тізімі, ашық түрде берілген *индекстеу сөздігі* немесе *дескрипторлық сөздік* қолданылады. Парадигматикалық қатынастары қамтылған дескрипторлық сөздік – *ақпарат іздестіру тезаурусы* (грек thesauros – *қор*) деп аталады.

Тезаурус арқылы іске асатын лексикалық бақылаудың мәні негізінен мына төмендегіге саяды:

1) синонимдік сипатты бақылау;

2) омонимдік сипаттың басын ашу;

3) бір сөзге ғана қатысты оның барлық түрлерін (варианттарын) нормалау, яғни олардың бәрінің басын біріктіру.

Лексикалық бақылауды құраушылардың әрбіреуіне қысқаша тоқталайық.

1. Синонимдерді бақылау әрекеті ақпаратты іздестіру жүйесіндегі лексикалық бақылаудың ең негізгі қызметі деуге болады. Мұндай әрекеттің барлық синонимдері тек бір ғана лексикалық вариантпен ауыстырылады. Бұл вариант *дескриптор* деп аталады және дескрипторлық тілдің лексикалық бірлігі орнына жүреді. Дескриптор жиынтығы және оған мағыналық жағынан жақын жатқан сөздер шартты эквиваленттіліктің тобын құрайды. Синонимдер қатарына мыналар қосылады:

- толық синонимдер (радар – радиолакатор, ЭВМ – компьютер);
- сөздердің фонетикалық (графикалық) варианттары;
- аббревиатуралар (ақпарат іздестіру жүйесі);

- мән-мағыналық (смысл) жағынан бір ғана мағынасымен сәйкес келетін сөздер (компьютер – машина, мақала – жұмыс).

2. Лексикалық бақылау омонимдерді (омографтарды) және көпмағыналы сөздерді ажыратуды да өз міндетіне алады. Мұндай мақсат оларға әртүрлі кодтарды (шартты белгілерді) сәйкестендіру арқылы іске асады. Ақпарат іздестіру тілінде лексикалық бірлік сандық (цифрлық) әліпби арқылы жазылады немесе таңбалар (белгілер) жүйесі арқылы беріледі, мысалы: *конденсаторлар* (бу үшін) – *конденсаторлар* (электрлік), *литье* (предмет) – *литье* (процесс), *меркурий* (металл) – *Меркурий* (планета). Кейбір зерттеулердің нәтижелері бойынша омонимдерді ажыратпау ақпараттық жағынан көп кедергі (информационный шум) келтірмейді.

3. Мәтінде кездесетін сөзтұлғалар тірек сөздер рөлін ойнап, индекстеу процесінде осы тірек сөздің (дескриптордың) сөздік формадағы нормаланған түріне (мысалы, атау септігі, жекеше) ауыстырылады. Кейде шартты эквиваленттіліктің бір ғана класына бір лексеманың грамматикалық формасына сайып қоймай, сонымен бірге әртүрлі негізді, тіпті әртүрлі сөз таптарына қатысты лексемалар ортақ лексикалық мағына негізінде (гиперлексема) біріктіріледі. Мысалы, *сверло* (тескіш) – *сверление* (тесу) және т.б.

Тезаурустың лексикалық бақылаудан кейінгі екінші маңызды қызметі – ақпарат іздестіру жүйесінің нәтижелігін, яғни толықтығын және дәлдігін арттыру. Бұған тек парадигматикалық және синтагматикалық қатынастарды қолдану арқылы қол жеткізуге болады.

Мәтіндегі лексикалық бірліктер арасындағы байланысты бейнелендіретін синтагматикалық қатынастарды белгілеу үшін арнайы грамматикалық құралдар, яғни рөлдер мен байланыс көрсеткіштері енгізіледі. Рөлдер көрсеткіші – қайсыбір дискриптор мәтінде (және оған сәйкес құжатты іздестіру бейнесінде) қандай мағынаға ие екендігін анықтайды және грамматикалық категориялар рөлін атқарады. Мұндай категориялардың мысалы ретінде "процесті", "қасиетті", "материяны", "ортаны" атауға болады. Көрсеткіштер құжатты іздестіру бейнесіндегі сәйкес дескрипторларға тіркеп жазылатын кодтар арқылы белгіленеді.

Құжаттың іздестіру бейнесіндегі дескрипторлардың қайсылары өзара байланыста (логикалық немесе синтаксистік) болатындығын байланыс көрсеткіштері анықтайды. Олар да құжаттың іздестіру бейнесіндегі дескрипторларға тіркеп жазылатын айрықша символдармен белгіленген. Рөл мен байланыс көрсеткіштерін қолдану дәлдік дәрежесін арттыруға мүмкіндік туғызады. Рөл мен байланыс көрсеткіштерін пайдалану арқылы құжаттарды индекстеу қосымша еңбек сіңіруді қажет етеді. Одан кейін де неше түрлі қателер кездесуі мүмкін. Сондықтан, әдетте, рөл мен байланыс көрсеткіштері іс жүзінде қолданылмайды, ал олардың орны басқа тәсілдермен толықтырылады.

Парадигматикалық (базистік) қатынастар. Мұндай қатынастар бойынша ақпарат іздестіру тіліндегі лексикалық бірліктер арасындағы мән-мағыналар (смысл) байланысы көрініс табады. Аталған байланыс мәнмәтінге қатыссыз, ақиқат әлеміндегі нысандар арасындағы қатынастарға тәуелді болады. Бұларға иерархиялық қатынастармен бірге көптеген ассоциативті қатынастарды (бөлік – бүтін, зат – қасиет, процесс – нәтиже және т.б.) жатқызуға болады. Мән-мағыналық (логикалық) байланыстар нақты түрде (эксплицитті) тезаурустарда белгіленеді. Ақпараттық іздестіру тезаурусы – белгілі білім саласының мазмұндық тұрғыдағы семантикалық моделі.

Ақпараттық іздестіру тезаурусын құрастыру аса көлемді және күрделі жұмыстың нәтижесі деуге болады және оның негізгі кезеңдеріне жататындар:

- ақпараттық іздестіру тезаурусының тақырыптық аумағын қамтуды айқындау;
- мәтіндік деректерді жинастыру;
- тірек лексикалық бірліктерді жинастыру (сөзтізбені қалыптастыру).
- лексикалық бірліктерді шартты эквиваленттік кластарға топтау;
- лексикалық бірліктің араларындағы парадигматикалық қатынастарды анықтау және олардың типтерін таңдау;
- сөздік мақаланы және көрсеткіштерді (нұсқағыштарды) құру;
- ақпараттық іздестіру тезаурусын қалыптастыру;
- экспертиза мен тіркеу жұмыстарын орындау.

Ақпараттық іздестіру тезаурусын құрастыру кезінде қосымша құрал ретінде, әдетте, ЭЕМ қолданылады. Оның құрастыру ережесі, құрамы мен түрпаты мемлекеттік стандартпен анықталады. Тезаурусты қалыптастыруда тұтынушы сұранысын қанағаттандыруға аса көңіл бөлінуі қажет. Оның түрпаттық сипаты мағыналық жағынан қажетті лингвистикалық бірлікті және барлық оған қатысты лексикалық бірліктерді тауып алуды қамтамасыз ету көзделеді. Әдетте ақпараттық іздестіру тезаурусы екі бөліктен тұрады:

1. Дескрипторлардың әліпбилік тізімі және оның сөздік мақаласымен (ұясымен) қоса есептелетін сөздік бөлігі, басқаша айтқанда, ол тезаурустың өзі деуге де болады. Сөздік мақала дескриптордан басқаша, әдетте, ол төрт ішкі жиыннан тұрады:
- тірек синоним-сөздер;
- "жоғарыда орналасқан" ("вышестоящие") дескрипторлар (осы дескриптормен "тек-түр" қатынасында тұратын).
- "төменде орналасқан" (нижестоящее) дескрипторлар (осы дескриптормен "тек-түр" қатынасында тұратын).
- ассоциативтік (ассоциацияланған) дескрипторлар.

2. Тезауруспен іздестіруді жеңілдететін әртүрлі көрсеткіштер (барлық дерек сөздердің әліпбилік көрсеткіші, сөз тіркестерінің элементтері үшін пермутациялық көрсеткіш, жиіліктік көрсеткіш және т.б.).

Тезаурусты құрастыру жұмысында қиыншылық тудыратын жеке бір мәселе, ол – ендірілетін сөз тіркестерін таңдау.

Индекстеу процесі, яғни қол күші әрекетімен орындалатын құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесіне енгізілетін дескрипторларды таңдау – *индекстеу әдістемесі* деп аталатын арнайы нұсқаулықпен анықталады. Солай бола тұра, индекстеу аса тапқырлықты талап етеді. Іздестірудің нәтижелігі көптеген жағдайда индекстеудің сапалығына тығыз байланысты. Ал оның өз тарапынан, ақпараттық іздестіру тезаурусының және индекстеудің әдістеменің сапалы болуына тікелей қатысы бар. Индекстеудің сапасын бағалауда іздестіру әрекетіндегідей индекстеудің толықтық (тереңдік) және дәлдік (егжей-тегжейлік) көрсеткіштері қолданылады. Ең алдымен, индекстеудің толықтық дәрежесі құжаттың іздестіру бейнесіне енгізілген дескрипторлар санымен, ал индекстеу дәлдігі – олардың "сапасымен", дәлірек айтқанда, таңдалып алынған дескрипторлардың негізгі құжаттың (сұратудың) мазмұнына жақындық дәрежесімен анықталады. Бұл сипаттамалар өзара кері тәуелділіктегі қатынаста тұрады. Индекстеудің толықтық дәрежесі мен іздестіру жұмысын қамтамасыз ету үшін **молынан индекстеу** (избыточное индексирование) тәсілі қолданылады. *Молынан индекстеу* ұғымын құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесін негізгілерімен мағыналық жағынан тәуелді болып келетін дескрипторлармен толықтыру деп түсіну қажет.

Ірі көлемді құжаттар базасын құрастыру, ақпаратпен алмасу қажеттігі және іздестіру нәтижелігін арттыру мәселелеріне байланысты индекстеу әрекетін

автоматтандыру проблемасы айрықша актуалдылыққа ие болды. Индекстеу процесін автоматтандыру мәселесін зерттеу ең алғаш Г.Лун (H.Luhn) және П.Баксендейл (P. Baxendale) және ресейлік Д.Г.Лахути есімдерімен тығыз байланысты.

Қазіргі кезде іске қосылған индекстеуді автоматтандыру жүйелері сапалық жағынан осыған ұқсас келетін автоматсыз түрде жасалатын интеллектуалды индекстеу жүйелерінен ешбір кем түспейді.

Жалпылама жағдайда, индекстеуді автоматтандыру сатысында сыртқы құжаттардан лингвистикалық бірліктерді таңдап алып, оны құжаттың іздестіру бейнесіне енгізу мақсатымен нормалау процесі жүргізіледі. Лексикалық бірлікті таңдап алу "оң сипатты" (мағыналы сөздер сөзтізбесі) немесе "теріс сипатты" (мағынасыз сөздер тізімі, "стоп"-сөз) сөздіктері бойынша орындалады.

Индекстерді автоматтандыру жүйесіндегі лексикалық бірлікті нормалау әрекеті морфологиялық талдау әдістері негізінде тілдік бірліктерді канондық түрге келтіру арқылы іске асады. Сонымен бірге, сөзтұлғаларды нормалау әдісінің тікелей іздестіру әрекеті кезінде де қолданылатын түрі бар. Нормалаудың ең көп тараған түрі (механизмі) – *кию әдісі* (немесе маскілеу). Бұл әдіс сұрату кезінде іздестіру термині ретінде тілдік бірліктің тек мағыналы бөлігі (негіз немесе квазинегіз) көрсетілетін жағдайда ғана орын алады.

Сапалы түрде индекстеуге қатысты қиындықтар тікелей рефераттар мен библиографиялық баяндаулар (описание) мәтіндері бойынша қажетті деген ақпаратты тауып алуды қамтамасыз ететін ақпараттық іздестіру жүйелерінің өмірге келуін тездетті. Ақпаратты іздестірудегі байқалатын ұстаным, ол іздестіру жүйесінің шектеулі табиғи тілге негізделген түрінің кең таралуы. Оларды тезауруссыз типті жүйе деп те атайды. Тезауруссыз ақпарат іздестіру жүйесінің кластары басқа жүйелерге қарағанда мынадай екі түрлі ерекшелікпен айырым табады:

- лексикалық бақылаудан бас тарту;
- құжаттарды индекстеуді автоматтандыру.

Мұндай ақпарат іздестіру жүйесінде координаттық индекстеу үшін табиғи тілдің сөздері, сөз тіркестері қолданылады. Шындығында, бұл жерде біз парадигматикасы мен сөздігі тұрақты түрде белгіленбеген дескрипторлық ақпаратты іздестіру тілімен әрекет етеміз. Ғалым А.И.Черныйдың пайымдауы бойынша, бұл жағдайда ақпарат іздестіру тілінің екі варианты қолданылады: біріншісі – табиғи тілдің тірек сөздері мен сөз тіркестерінің нөлдік қалпы, ал екіншісі – жоғары дәрежелі логикалық өңдеу (сұратуды молынан индекстеу үшін) (*Черный, 1975*).

Аталған жүйелер УНИТЕРМ жүйесінен бастама алады. Мұндай тезаурустың қызметі мен қолданудағы ұтымдылығы жайлы жоғарыдағы баяндауымызда қысқаша айтылған болатын. Бірақ лексиканы бақыламайтын жүйелердің де өзіне тән құндылығы бар. Ең алдымен, бұған олардың аса дәлдік сипатын жатқызуға болады. Ешбір тезаурус табиғи тілдің лексикалық әралуандығын толығымен қамти алмайтындығы ақиқат шындық екені белгілі. Сондықтан индекстеу процесі құжат мазмұнын толық ашпайды, ал, тіпті, кейбір жағдайларда мазмұн өрескел бұрмаланады. Әдетте, көптеген жүйелерде тар сипаттағы көпаспектілі сұратулар басым болып келеді. Сол себепті табиғи тілге бейімделген жүйелердің үлкен артықшылығы бар, себебі, олар толық түрдегі ерекшелік сипатқа ие. Ал тезаурус ешқашанда ондай ерекшелікке ие бола алмайды. Сонымен бірге, әдетте, әртүрлі ақпараттық-іздестіру тезаурустары да бір-бірімен сәйкес келмейді. Кезінде ондай тезаурустардың өзара сәйкес келу проблемасы ақпарат іздестіру жүйесі үшін негізгі мәселе болғанын да айта кетуге болады. Бұл мәселенің шешімін табу үшін әртүрлі тәсілдер қолданыс тапты, солардың ішінен аралық тілді құрастыру тәсілі де орын

алған болатын. Бірақ бұл мәселенің шешімі табылмай, бұл проблеманы табиғи тілге бейімделген жүйелер толығымен күн тәртібінен алып тасталды.

Ақпарат іздестіру жүйелеріне баға беруге байланысты жүргізілген тәжірибелердің нәтижелеріне қарағанда, мәтін арқылы іздестіру жүйелері өзінің көптеген көрсеткіштерімен тезаурус типті жүйелерден қалыспайтындығын байқатты. Негізінен, ондай көрсеткіштерге дәлдік дәрежесінің ұлғаюы мен толықтық сипатының төмендеуі тән деуге болады. Бұл заңды да. Табиғи тіл арқылы іздестірудің ең жақсы нәтижелеріне жету үшін табиғи тілге тән "кемшілік" жақтарының орнын толтыру керек болды. Ол үшін ақпарат іздестіру жүйені екі түрлі мүмкіндіктердің (механизмдердің) жиынтығы деп қарастыруға болады: мағына ажырату және мағынаны сәйкестендіру. Бұл механизмдердің "күштілік" деңгейі бір-біріне сәйкес келулері керек. Айталық, егер біз ақпарат іздестіру жүйесінің мағына ажырату күштілігін арттыратын болсақ, міндетті түрде мағынаның сәйкестік механизмінің күшін де ұлғайта түсуіміз керек болады. Лексиканы нормалау мен лексикалық бақылау әрекеттері мағынаны сәйкестендіру қызметін жеңілдетеді.

Ақпарат іздестіру жүйесіне құжаттардың (рефераттардың) толық мәтінін енгізу құжаттың барлық лексикалық бірліктері мен олардың арасындағы мәнмәтіндік қатынастарды толық сақтап қалуға мүмкіндік туғызады. Байланыста тұратын мәтіндер бойынша әрекет ететін іздестіру жүйелерінің дәстүрлі дескрипторлық іздестіруден айырмашылығы – әрбір сөзтұлғаның мәтіндегі дәл орнын көрсету мүмкіндігі. Яғни бұл жердегі "орын" деп отырғанымыз сол сөзтұлға кездескен абзацтың нөмірі, сөйлемнің нөмірі, сөйлем ішіндегі сөздің нөмірі және т.б. Сөзтұлғалардың мәтін ішіндегі дәл орнын ("адресін") білу сұрату мен іздестіру алгоритмінің арнайы мәнмәтіндік операторлар тілін туындатты. Аталған операторлар тілі сөзтұлғалар арақашықтығын және басқа да позициялық сипаттамаларға қатысты деректерді ("позициялық операторлар") пайдаланады. Ал бұл жағдай лингвистикалық бірліктердің синтагматикалық қатынастарын барынша жеңіл және икемді тәсілмен қолдануға мүмкіндік туғызады. Мұндай ақпарат іздестіру жүйесі – сызықтық грамматикалық жүйе деп аталады. Ондай жүйелерге ДИАЛОГ, CDS/ISIS, СКОБКИ, РАСПРИ деп аталатын *қолданбалы программалар пакеттерін* мысал ретінде келтіруге болады. Позициялық операторларды диалогтық іздестіру тәртібімен бірге қолдану жолы қарапайым әдіс-тәсілдер арқылы терминдердің жалған координатымен шатастырып алмауға мүмкіндік туғызады. Ал мұндай жалған сөзнұсқағыштық жағдай үлкен көлемді мәтіндер бойынша іздестіру жұмыстарын жүргізу кезінде сөзсіз болады.

Тезауруссыз ақпарат іздестіру жүйелерінде іздестіру ұйғарымын құрастыру сатысында мағынаны сәйкестендіру (және толықтықты арттыру) проблемасы өз шешімін табады. Себебі, мұндай ақпарат іздестіру жүйелерінде сұратуларды молынан индекстеу жұмысы іске асады. Онда іздестіру ұйғарымындағы әрбір терминге синонимдік немесе мағына жағынан жақын жатқан терминдер тіркеліп жазылып, ақпараттық іздестіру тезаурусындағы сөздік мақаланың толық пішіні орын алады.

Ақпарат іздестіру жүйесін лингвистикалық жабдықтаудың маңызды құрамдық бөлігі ретінде әрқилы сөздіктерді атауға болады. Олар автоматты немесе автоматсыз индекстеу тұсында, құжаттар мен сұратуларды жүйеге енгізу кезінде орын алатын қателерді автоматты түрде тауып алуда және т.б. жағдайларда қолданыс табады.

Мәтіндердегі қателерді автоматты түрде тауып алу әдістемесін екі үлкен топқа бөліп қарастыруға болады. Олар: бақылаудың полиграмдық (көпграфемалық) әдістері және сөздік арқылы бақылау әдістері. Бірінші жағдайда, қатені тауып алу құралы ретінде қарастырып отырған тілде мүмкін болатын екіәріптік немесе үшәріптік

тіркестер алынады. Екінші жағдайда мәтін бойындағы сөздерге жасалған морфологиялық талдаудың нәтижесі эталон ретінде алынған сөздікпен салыстырылады. Соңғы әдістің, яғни сөздік әдісі бойынша іздестіру мен қателерді өңдеу жұмыстарының тиімді екендігі тәжірибе жүзінде дәлелденіп отыр. Бірақ жұмыстың нәтижелілігі сөздіктің көлемі мен талданатын мәтіндерді қамту толықтығына байланысты болып келеді. Қатені табуға арналған ішкі жүйелердегі негізгі орын сөзөзгерту мен сөзжасам құбылыстарын ескере алатын автоматтанған морфологиялық әрекеттерге беріледі.

Ақпарат іздестіру жүйесін лингвистикалық жабдықтау жекелік сипаттағы маңызды бөлігі – машиналық сөздіктерді құрастыру мен іске қосуға қажетті құралдар мен тәсілдер. Әдетте мұндай мәселелер біріктіріліп, автоматтанған сөздік қызметі ұғымына енеді.

Шығарылу критерийі. Құжат пен сұратуды бір-біріне сәйкестендіргенде құжаттың сұратуға қарағанда релевантты болатындығын анықтау қажет болады. Осындай әрекеттің нәтижесінде сұрату бойынша тауып алынған құжаттың *шығарылуына* немесе *шығарылмауына* қатысты шешім қабылдау керек. Мұндай шешім – шығарылу критерийі арқылы қабылданады. Шығарылу критерийі басқаша атпен *мағыналық сәйкестік критерийі* деп те аталады. Сонымен, шығарылу критерийі немесе мағыналық сәйкестік критерийі дегеніміз құжат пен сұратудың формалды релевантты дәрежесін (яғни құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесі сәйкестігін) анықтайтын формалды ережелер деуге болады.

Дескрипторлық ақпарат іздестіру жүйесінде қолданылатын барлық шығарылу критерийлерін екі үлкен топқа бөліп қарастыруға болады. Олар: *есептелінетін* (вычисляемые) және *логикалық*.

Есептелінетін шығарылу критерийі сандық параметрлер бойынша әрекет етеді. Бұған құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің көлемге қатысты сипаттамалары да қамтылады. Бірқатар критерийлер олардың терминдер жиындары қиылысуларының теоретикалық-жиындық амалдарына негізделеді. Мұндай қиылысу шамасы пайыздық мөлшерде анықталуы мүмкін. Ең қарапайым жағдай – құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің толық сәйкес келуі. Бірақ мұндай сәйкестік аса қатал шарт түріне қатысты. Әдетте, тәжірибеде қиылысу шамасының қайсыбір "шекаралық" (пороговое) шамасы (мысалы, 50%) беріледі. Ал бұл шамадан жоғары сатыда тек релеванттылықтың деңгейі ғана тұрады. Шығарылу критерийінің қарапайым түрі құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесіне енетін терминдердің бірдей дәрежедегі маңыздылығына келіп саяды. Бұл жағдайда шығарылу критерийінің шамасы мынадай өрнекпен (формуламен) анықталады:

$$R = \frac{N_i}{N_z} \cdot 100\%$$

мұндағы N_i – құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесіндегі терминдерді іздестіру кезіндегі сәйкестік саны; N_z – сұратудың іздестіру бейнесіндегі терминдердің жалпы саны.

Сонымен қатар, көпшілік жағдайда әртүрлі терминдер, әсіресе іздестіру ұйғарымында, бірмәнді бола бермейді. Осыған байланысты терминге қатысты "салмақтылық" немесе "салмақ" ұғымы туындайды. Терминдердің салмағын анықтау амалы терминдерді "өлшеу" деп аталып жүр. Өлшеу тәсілдері әртүрлі болуы мүмкін. Олардың ең қарапайым түрі келесі пайымдауға негізделеді. Іздестіру

ұйғарымын құрастыруда тұтынушы әрбір іздестіру ұйғарымындағы дескрипторды белгілі санмен "бағалайды". Міне, осы сан – *салмақ коэффициенті* деп аталады. Ол осымен қатар, тұтынушы салмақтар қосындысының "табалдырықтық" мәнін алдын ала көрсете алады. Сонымен бірге терминдерді "өлшеудің" автоматтанған әдістері де қолданыс табады. Бұл әдіске құжаттарды автоматты түрде индекстеу әдісін де жатқызуға болады. Мысалы, қайсыбір сөздің қайталанып қолданылуының қатынастық жиілігі қарастырылуы мүмкін. Осы құжаттағы сөздердің қатынастық жиіліктері барлық құжаттағы осындай сөздердің қатынастық жиілігінен асып түсу мүмкіндігі құжаттың негізгі мазмұны тұрғысынан аса маңызды деп саналады. Ал келесі бір тәсіл бойынша, құжаттың іздестіру бейнесіне қосу мақсатымен құжаттың ең маңызды деген тұстарынан (тақырыбы, түйіні және т.б.) таңдалып алынған терминдерге жоғары салмақтар тіркеліп жазылады.

Құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің сәйкестік дәрежесін анықтайтын шығарылу критерийі әртүрлі тәсілдермен есептеліп шығарылуы мүмкін. Жалпы жағдайда, әрбір "құжат – сұрату" жұптары үшін векторларды салыстыру әдісінің көмегімен ұқсастық коэффициентінің шамасы есептеліп шығарылады. Осының нәтижесінде шығарылу құжаттарын, олардың ұқсастық коэффициенттерінің кему тәртібімен реттеуге және шығарылатын құжаттардың санын айқындауға мүмкіндік жасалады. Құжаттың саны алдын ала берілген "табалдырықтық" мәнге байланысты болады.

Қазіргі кездегі көптеген тәжірибелердің нәтижелері негізінде *іздестіру коэффициенті* қолданыс табатын барлық ақпарат іздестіру жүйелерінің жоғары тиімділігі дәлелденіп отыр. Бірақ іс жүзінде логикалық коэффициентке негізделген ақпарат іздестіру жүйелері іздестіру коэффициентіне негізделген жүйелерді бірте-бірте ығыстыра бастады. Мұндай жүйелер сұратудың іздестіру бейнесі логикалық – "И", "ИЛИ", "НЕ" операторлары арқылы құрылады (орыс тілінің мағынасындағы сөздер &, V, ¬ таңбаларымен белгіленеді). Осының нәтижесінде бір ғана ақпарат іздестіру жүйесінің шеңберінде екі түрлі ақпарат іздестіру тілі жайлы сөз етуге мүмкіндік туады. Оның екіншісін *булелік алгебраға* (булевая алгебра) негізделген сұратудың арнайы тілі деп ұғыну керек. Мұндай жағдайда, егер сөз құжаттың іздестіру бейнесінде бар болса, тірек сөздер немесе іздестіру ұйғарымының дескрипторлары "1" ("шындық") мәнін, ал ол сөз құжаттың іздестіру бейнесінде жоқ болса "0" (нөль – "жалған") мәнін қабылдайды. Егер сұратудың логикалық формуласы, толығымен алғанда, осы құжатқа қатысты "шындық" мәніне тең болса, онда табылған құжат сұратуға релевантты, яғни олардың мазмұндары бір-біріне сәйкес келеді деп есептеледі.

Көпшілік жағдайда іздестіру ұйғарымы конъюнктивтік нормадағы формалық (дизъюнкциялардың конъюнкциясы) түрде жазылады. Әдетте, синонимиялық қатынаста тұратын терминдерді "ИЛИ" (немесе) (дезъюнкция) операторлары арқылы байланыстырады. Ал "И" (мен) (конъюнкция) операторы арқылы сұратудың әртүрлі аспектілерін анықтайтын терминдерді (немесе топталған терминдерді) өзара байланыста ұстайды. **Булелік операторларды** қолдану құжаттар мен сұратуларды салыстырудың тұтынушыға түсінікті пішіндегі логикалық қисындылығын қамтамасыз етеді.

Сонымен бірге, логикалық шығарылу критерийі ескерілетін ақпарат іздестіру жүйелерде тұтынушыға ұсынылатын құжаттарды маңыздылығына қарай реттеу мен олардың көлемін бақылау қиындық туғызуда. Сондықтан тұтынушының өз қалағанындай нәтижеге қол жеткізе алмауына байланысты ақпарат іздестіру жүйесіндегі шығарылудың логикалық критерийінің "жеңісін" тек уақытша деп санауға болады.

Ақпараттық жабдықтау. Ақпараттық іздестіру жүйелерінде бірінші және екінші құжаттық "нұсқалар" ажыратылады. Бірінші "нұсқа" ретінде ЭЕМ жадында сақталатын іздестіру жиымдары (құжаттың іздестіру бейнесі, сұратудың іздестіру бейнесі) түсінік табады. Екінші "нұсқа" – әдетте, ЭЕМ-нан тыс (қағазда немесе микротаспаларда) сақталатын алғашқы құжаттардың жиымдары. Бірақ қазіргі кездегі ЭЕМ-дардың жадының ұлғаю сипаты толықмәтіндік деректер базалары санының көбеюіне әкеліп соқты да, соның нәтижесінде бірінші және екінші нұсқалық құжаттар бірігіп, біртұтастық көрініс тапты.

Қазіргі заманға сай ақпарат іздестіру жүйелері кең түрдегі ақпарат алмасу жиымдарымен (деректер базасымен) әрекет етуге арналатыны белгілі. Сондықтан олардың электрондық тұрпатының көрінісі арнайы халықаралық және мемлекеттік стандарттармен айқындалады, яғни басқаша айтқанда, олар тұрақты ережеге бағындырылады. Қазіргі кездегі ең кең тараған халықаралық стандарт – ISO 2709 (ГОСТ 7.14-84). Аталған стандартқа сәйкес жазылған барлық құжаттар мынадай бірыңғай құрылымға ие:

- жазу маркері – жазба бастамасында орналасқан және барлық жазбаға қатысты параметрлерді бойында сақтайтын, алдын ала белгіленген тұрақты ұзындықтағы өріс;
- анықтағыш – құжатта көрініс табатын деректердің әрбір библиографиялық бөліктеріне қатысты мазмұн элементтерінің жиынтығы;
- деректер элементтерін бойында ұстайтын деректер аясы (область данных).

Әрбір құжаттың толығуы (деректердің библиографиялық бөліктерінің атаулары, олардың семантикасы, берілу тұрпаты) арнайы унификациялауды, яғни бірізге келтіруді қажет етеді. Мұндай бірізділіктердің кең тараған түрі MARC ұясына жататындар (USMARC, UNIMARK және т.б.)

Техникалық және программалық жабдықтау. Техникалық және программалық жабдықтау – іздестіру қызметін (функциясын) іске асыру құралдары. Бұл жайтқа біз тек жалпылама түрде ғана тоқталамыз.

Ең алғашқы ақпарат іздестіру жүйелері адамның қатынасымен әрекет ететін және механикаландырылған жүйелер болатын. Қазіргі заманғы ақпарат іздестіру жүйелері ЭЕМ-ның негізінде іске асады. Жүйенің құрамындағы техникалық құралдарға қойылатын негізгі талап – олардың тез әрекет етуі мен жадының көлемділігі. Соңғы жылдары электрондық техниканың мүмкіндіктерінің барынша артуы көпшілікке мәлім, яғни қазіргі заманғы ЭЕМ-ның әрекет ету жылдамдығы (секундына миллиондаған амалдар атқарады) магнитті және оптикалық дискілер жадының көлемі *мега, гига* және *терабайттармен* өлшенеді. Практиалық тұрғыда қарастырғанда ЭЕМ-ның мұндай сипаттамалары кез келген іздестіру жүйесін құруға мүмкіндік туғызады. Ақпараттың шығарылуын қамтамасыз ететін неше түрлі "көру" мүмкіндігі бар терминалдық құралдар және әркилы қағаз бетіне шығаратын құралдар (принтерлер) тез басу мүмкіндігімен бірге таңдау мүмкіндігіне де ие. Қазіргі ақпарат іздестіру жүйелері, әдетте, байланыс желілерінің және программалық-аппараттық комплекстердің негізінде ақпаратты қашықтықтан іздестіруді де орындай алады. Ақпараттың көлеміне және ақпарат іздестіру жүйенің мақсатына қарай іздестіру жүйелері ірі әмбебапты ЭЕМ-дар мен персоналды компьютерлер негізінде де құрастырылуда. Қажеттікке сай, персоналды ЭЕМ-дар (компьютерлер) үлестірілген базалар мен деректер банкілерін құруға мүмкіндік туындататын локалды желілерге біріктіріледі.

Ақпарат іздестіру жүйенің қызметін сүйемелдейтін программалық жабдықтауды шартты түрде екіге бөліп қарастыруға болады: *жалпы жүйелік программалық*

жабдықтау және *қолданбалы программалық жабдықтау*. Қолданбалы программалық жабдықтау нақты ақпарат іздестіру жүйесінің барлық қызметін іске асырады: ақпаратты енгізу, бақылау, сақтау, индекстеу, іздестіру, шығару. Программалық құралдардың құрамы және оларды іске асыратын әдістер көпшілік жағдайда лингвистикалық және ақпараттық жабдықтауға тәуелді болып келеді.

Ақпарат іздестіру жүйесінің программалық жабдықтау мәселесін зерттеуде деректер моделі мен ақпараттық базаны (іздестіру жиымын) құрастыру мәселелері маңызды орын алады. Деректерді ұйымдастыру тәсілін таңдау әртүрлі факторларға тәуелді болып келеді. Мысалы, деректердің көлеміне, оларды қолданысқа енгізу жиілігіне, өзгерістердің сипаты мен динамикасына, техникалық құралдардың мүмкіндігіне, іздестіру режиміне (ИРИ, "ретро"), іздестіру логикасы мен шығарылу критерийіне. Деректерді ұйымдастыру тәсілдері ішінен деректерге еркін "кіру" әдістерінің бірізді және төтенше түрлері айырым табады. Бірізділікпен енудің (кірудің) негізінде, ол берілістерді біртіндеп, яғни бірінен кейін бірін алып өңдеу мүмкіндігі орындалады. Ал төтенше әдіс бойынша деректерге "қол жеткізу" үшін арнайы алгоритмдер құрастырылады және әдетте еркін таңдалып алынған кез келген деректі тез табуға мүмкіндік жасайтын қосымша файлдар (деректер жиыны) құрастырылады.

Ақпарат іздестіру жүйесіне қатысты программалық жабдықтау мәселесін зерттеп, іске қосу үшін алгоритмдік тілдер мен операционалдық амалдар жүйелерінің құрамына енетін арнайы құралдар қолданылады. Іс жүзінде негізінен арнайы міндет жүктелген стандартты программалық комплекс қолданылады. Дәлірек айтқанда, құжатты ақпараттық іздестіруге арналған *қолданбалы программалар пакеттері* қолданылады. Мысал ретінде Ресей кеңістігінде көп тараған ППП ДИАЛОГ және COS/ISIS қолданбалы программалар пакетін атауға болады. Бұл пакеттер, әдетте, іздестіру жүйесінің өзегін құруға мүмкіндік жасайды және мынадай құрамнан тұрады:

- деректер базасын құруға қажетті құралдар;
- деректер базасын толықтыру мен оған түзетулер енгізуге мүмкіндік жасайтын құралдар;
- сұрату мен іздестіруді құрастыруға қажетті құралдар;
- тұтынушы мен деректер базасы аралығындағы диалогтық қарым-қатынасты іске асыратын құралдар;
- ақпаратты шығару құралдары.

Аталған пакеттер тікелей "кіру" әдісі бойынша тез іздестіру үшін инверттелінген (дағдылы тәртібі өзгерген) файлдарды (индекстерді) құруға мүмкіндік береді. Инверттелінген файлдарды берілістің (деректердің) кез келген элементі бойынша құрастыруға болады. Ондай элементтерге мәтіндік өрістің сөзтұлғаларының да қатысы бар. Мәтіннің кез келген сөзі сол мәтіндегі "мекенжайымен" ("адресі") бірге инверттелген файлда тіркеледі. Осының нәтижесінде, іздестіру кезінде сұратудағы терминдердің мәтін бөлігінің қажетті деген жеріне (мысалы, сөйлемге) дәлме-дәл түсуі үшін мәнмәтіндік (позициялық) операторларды қолдануға мүмкіндік туады. Бұл әсіресе толық мәтіндер үшін маңызды деп саналады. Ең бір дамыған толықмәтіндік ақпарат іздестіру жүйесі ретінде АРТЕФАКТ жүйесін атауға болады. Бұл жүйеде сөзтұлғаларды морфологиялық нормалау, деректерді "қысу" және сұратудың дамыған тілі сияқты әрекеттерді қолдану іске асады. Мәнмәтіндік операторлар іздестірудің сапалылығы мен икемділігін арттыруға маңызды сұратудағы сөз тіркестерін құруға мүмкіндік жасайды. Іздестіру мүмкіндігін арттыру үшін форматтық өріс бойынша іздестіру механизмін іске қосуға болады. Бұл тәсіл

бойынша арифметикалық салыстыру операторлары (үлкен, кіші, тең және т.б.) арқылы қосымша шарттарды енгізуге жағдай жасалады.

Белгілі шектеуліктерді ескере отырып, форматтық өріс бойынша орындалатын іздестіру механизмі фактографиялық іздестіру кезінде де қолданылуы мүмкін. Солай бола тұра, фактографиялық ақпараттық іздестіру жүйесін, кейбір жағдайларда құжаттық іздестіру жүйесін де құру үшін *деректер базасын басқару жүйесі* қолданылады.

Автоматтанған фактографиялық ақпараттық іздестіру жүйесі – нақты деректердің (фактілердің), белгілі нысандардың қасиеттері және т.б. жайлы ақпараттарды сақтайтын ақпараттық жүйе. Автоматтанған фактографиялық ақпараттық іздестіру жүйені сипаттауда, әдетте, деректер базасы теориясының терминологиясы қолданылады. Себебі, ол жеткілікті түрде дамыған және автоматтанған фактографиялық ақпараттық іздестіру жүйені іске қосудың практикалық жағы қазіргі кезде осы теорияның негізінде іске асуда. Деректер базасы ұғымын ЭЕМ жадында сақталатын деректер мен олардың арасындағы қатынастардың жиынтығы деп түсіну қажет. Белгілі бір мағынада деректер базасы "ақиқат болмыстың үзіндісінің" моделі немесе оның ақпараттық бейнесі деуге болады. Ал деректер базасында сақталатын құбылыстар, деректер (фактілер), заттар және т.б. жайлы ақпараттар нысан (объект) деп аталады. Басқа сөзбен айтқанда, нысан – біртұтас күйінде тұжырымдауға болатын деректер базасындағы ақиқаттықты бейнелейтін элемент. Мысалы, нысан ретінде ұшақтың рейсі (жолы), механизмнің бөлшегі, өндірістің қызметкері және т.б. алуға болады. Кез келген нысан әртүрлі қасиетке, сапаға ие болады. Міне, бұл аталғандарды нысанның *белгілері* немесе *атрибуттары* деп атайды. Атрибуттардың жиынтығы нысанды анықтайды да, ал ақпараттық көзқарас бойынша, нысан – осы атрибуттар жиыны деп саналады. Мәселен, *ұшақ рейсі* атты нысан мынадай атрибуттарды сипаттай алады: рейс нөмірі, ұшақтың типі, сапарға аттану пункті, сапардың аяқталу пункті, сапарға аттану күні, уақыты, сапардың аяқталу уақыты, ұшақтың әуеде болу уақыты, билет құны және т.б.

Егер программаларды жазу тілінің терминімен айтсақ, онда әрбір нысан бір "жазба" (запись) түрінде көрініс табады. Әрбір **жазба** атрибутқа сәйкес келетін өрістерден тұрады.

Осындай терминологияға сүйенетін ақпараттық жүйелерді кейде *нысанды-белгілік ақпараттық жүйе* деп те атайды. Сонымен, деректер базасы – белгілі бір нысандардың жиыны (атрибуттар жиыны) және олардың аралық байланысы.

Деректер базасын құрылымдық тұрғыда қарастырғанда әдетте оны үш түрге бөледі: *иерархиялық, желілік* және *реляциондық*.

Деректік базаның құрылымы "шежірелік" (бұтақ таралымы) принципіне негізделсе, оны *иерархиялық құрылымдық деректер базасы* деп атайды. Егер деректік базаның құрылымы еркін желі құрылымындай болса – *желілік құрылымдық деректер базасы* болмақ. *Иерархиялық құрылымдағы деректер базасына* өңдеу жұмыстарын іске асыру үшін жан-жақты зерттелген математикалық құралдар кең түрде қолданылады. Ал желілік құрылымдағы деректер базалармен әрекет ету соңғы кезде көптеген қиындықтарға соқтырып жүр. Сондықтан желілік деректер базасына қарағанда иерархиялық деректер базасы бойынша іздестіру жұмысын атқару аса нәтижелі болатындығы дәлелденді. Бірақ кез келген ақпаратты *иерархиялық құрылымда* бейнелеу ойдағыдай бола бермейді. Сол себепті егер ақпаратты баяндау тәсілі иерархиялық сипатта болса ғана, іздестіру, өңдеу жұмыстары иерархиялық деректер базасына қолданылуы керек. Соңғы кезде жиі қолданыс тауып жүрген реляциялық деректер базалар тікбұрыштық кестелердің жиыны түрінде көрініс табады. Ал кестелер арасындағы байланыс – айқыш сілтемелер арқылы іске асады.

Деректер базасы бойынша жұмыс істеу мүмкіндігіне ие болу үшін арнайы жазылған программалар қажет. Деректер базасындағы жұмысты қамтамасыз ететін программалар топтамасын *деректер базасын басқару жүйесі* деп атайды. Деректер базасын басқару жүйесінің құрамы әртүрлі деректер базалармен жұмыс істеуге мүмкіндік тудыратын барынша әмбебаптық сипаттағы программалардан тұруы қажет деп саналады.

Деректер базасын басқару жүйесі деректер базасының құрамдық жағын құрастыруды, деректер базасындағы ақпараттық өзгерісті (модификацияны), деректер базасындағы ақпаратты іздестіруді, тұтынушыға ақпараттың қолайлы түрде шығарылуы мен басқа да қосымша функцияларды қамтамасыз етуді көздейді. Осы аталғандарға толығырақ тоқталайық.

Деректер базасының құрамын құрастыру. Деректер базасын зерттеп, іске қосу барысында оның құрамдық жағына көп көңіл бөлінеді. Ақпаратты сақтау мен өңдеу нәтижелі болу үшін қандай ақпарат түрін сақтаудың қажеттігі, деректер базасында қандай нысандардың типтері орын алатындығы алдын ала айқындалуы қажет. Әрбір нысан типтері үшін оларды айқындайтын атрибуттар топтамасы анықталуы керек болады. Сонымен бірге әртүрлі типтегі нысандар арасында мүмкін болатын байланыс типтері де анықтауды қажет етеді. Барлық нысандардың типтері бейнеленген және әрбір нысан типінің атрибуттар топтамасы мен мүмкін болатын байланыс типтері көрініс тапқан сызбаны *жалпы концептуалдық сызба* немесе *деректер моделі* деп атайды.

Әдетте, деректер базасын зерттеу онда бейнеленетін ақпаратқа қатысты пәндік аяны зерттеуден және жалпы концептуалды сызбаны бейнелеуден басталады. Деректер базасын басқару жүйесінде жалпы концептуалдық сызбаның негізінде құрастырылатын деректер базасының құрылымы жайлы ақпаратты енгізу мүмкіндігі ойластырылуы керек. Осындай жағдайда әрбір атрибуттың деректер типін анықтау қажеттігі (программалау мағынасында) туындайды.

Әдетте деректер базасын құрастырғанда артықтық сипаттағы ақпараттың оның жадында сақталмағаны жөн деп саналады. Сондықтан, кейбір жағдайда, деректер базасын артық болмайтын (неизбыточная) деректер жиыны деп анықтайды. Бірақ іс жүзінде мұндай нәтижеге жету оңай емес. Көбінде ақпаратты іздестіру жұмысының ыңғайлы болуы үшін ақпараттың артықтығы қолайлы деп те саналады. Мысалы, *ұшақ рейсі* деп аталатын нысанның мынадай атрибуттары *ұшу уақыты, қону уақыты* және *әуеде болу уақыты* ішінен екеуі ғана белгілі болса, үшінші атрибутты әр уақытта есептеп шығаруға болады. Бірақ егер біз деректер базасынан ақпаратты іздестіру әрекетін *ұшақ рейсі* сипаттамаларының әрбіреуі бойынша жүргізу қажет деп санасақ, үш түрлі атрибуттың да орын алғаны жөн болады. Солай болғанның өзінде де, жалпылама жағдайда, ақпарат артықтығының аз мөлшерде болуына талпынуымыз қажет.

Деректер базасындағы ақпараттың артықтық сипатының қауіптілігі тек қосымша жад көлемін қажет етуде ғана емес, ең алдымен, деректер модификациясының өзгеріске ұшырау салдарынан қадағалауға қажетті қосымша бақылау әрекетінің қажеттігі болуында. Сондықтан зерттеушілердің тұжырымдауынша, деректер базасындағы артықтықтың тек бақылау жұмыстарына ғана қатысты болғаны жөн деп саналады.

Деректер базасындағы ақпаратқа өзгерістер (модификациялар) енгізу. Деректер базасын басқару жүйесінің деректер базасында орналасқан ақпаратқа тиісті өзгерістер енгізуге мүмкіндіктері болуы қажет. Ондай өзгерістердің типі мынадай болуы мүмкін: *жаңа жазбаны* (новая запись) немесе *жаңа нысанды қосу, бар жазбаны*

өшіру, бар жазбаның атрибут мәнін өзгерту. Деректер базасының құрылымын модификациялау ең күрделі өзгеріс түріне жатады. Іс жүзінде деректердің жалпы концептуалдық сызбасының идеалды түрін (өзгеріске жатпайтын) алдын ала болжамдау мүмкін бола бермейді. Сондықтан деректер базасын қолдану кезінде кейбір нысандар қажетсіз, ал қайсыбірінің жетіспеуі немесе артық болуы, тіпті кейбір тәуелділіктерді басқаша түрде ұйымдастыру қажет болады. Деректер базасын басқару жүйесінде деректер базасының құрылымын өзгерту тұсында ондағы қажетті деген ақпаратты сақтап қалу мүмкіндігі жасалуы қажет.

Ақпаратты іздестіру. Деректер базасын басқару жүйесі қызметінің ішінде ең бір маңызды деп саналатыны – тұтынушының сұранысы бойынша жүргізілген іздестіру әрекеттерінің нәтижелі болуын қамтамасыз ету. Іздестіру жұмысы дәл және тез уақыт аралығында орындалуы қажет деп ұйғарым жасалады. Ал ол үшін деректер жиынының белгілі пішінде ұйымдастырылуы қажет болады. Реттелген тізбе деректері бойынша іздестіру әрекетінің ойдағыдай нәтижелі болатындығы белгілі. Ал егер деректер тізбесі ретсіз түрде орналасса, онда қажетті деген нысанды тауып алу үшін барлық деректер жиынын толығымен, яғни басынан аяғына дейін қарастырып шығуға тура келеді. Мұндай әрекетке көп уақыт кетеді, ал уақытты үнемдеу үшін іздестіруді әртүрлі атрибуттардың мәні негізінде жүргізген жөн болады. Бұл мүмкіндік әдетте тек инверсияланған индекстік файлдарды құрастыру арқылы іске асады. Ондай файлдарда атрибуттардың мәні мен қажетті деген жазба түрлеріне сілтеме жасайтын арнайы түрдегі көрсеткіштер орын алады. Артықтық сипаттағы ақпаратты индекстік файлдар өз бойында сақтайды. Сондықтан мұндай ақпараттық артықтықты бақылау үшін деректер базасын басқару жүйесінде ақпаратты өзгертуге (модификациялауға) қажетті арнайы механизмнің орын алғаны жөн. Әрбір атрибутты олардың мәндері бойынша тез тауып алу үшін оның өзіне ғана сәйкес келетін, индекстік файлы болуы қажет. Яғни деректер базасын басқару жүйесінде ондай міндетті атқаратын арнайы программа (бағдарлама) құрастырылуы керек деген сөз. Мәні бойынша атрибутты тез іздестіру мүмкіндігін оның **"кілті"** немесе *кілттік атрибуты* деп атайды. Жазбалар файлының реттелуіне қатысты "кілтті" *бірінші реттегі кілт,* ал басқаларын *екінші реттегі кілттер* деп ажыратады. Әдетте *бірінші реттегі кілттің* өзіндік артықшылыққа ие болуына немесе жазбаның бірмағыналы теңестірулерге мүмкіндігі болуына талпыну керек болады. *Екінші реттегі кілттердің* мәндері әртүрлі жазбаларда қайталанып тұруы мүмкін. Сондықтан *бірінші реттегі кілт* бойынша іздестіруді ұйымдастырудың жылдамдығы аса жоғары, яғни ол максимум шамаға ие. Деректер базасын басқару жүйесі іздестіру жұмысын кілттік атрибуттардың мәні бойынша ғана емес, сонымен бірге кілтсіз атрибуттар негізінде де іске асыра алады. Бірақ соңғы жағдайда іздестіру жылдамдығы кемиді. Деректер базасында ақпаратты іздестіруге қатысты қарапайым екі түрлі типтегі сұратулар орын алады. Бірінші тип – айқын атрибуттың мәнін (шамасын) белгілі нысан бойынша (бірінші реттегі кілт) іздестіру. Сұратудың бірінші типіне мынадай мысалды келтіруге болар еді: *"Егер ұшақтың рейс нөмірі (бірінші реттегі кілт) белгілі болса, оның ұшу уақытын қалай табуға болады?".* Ал сұратудың екінші типі нысанды (немесе нысандарды) деп айқын атрибуттың белгілі мәні бойынша іздестіру әрекетін атайды. Мысалы: *"Егер ұшақтың қону пункті (жері) белгілі болса, оның рейс нөмірін (немесе рейс нөмірлерін) қалай табуға болады?".* Егер сұрату түрі күрделі болатын болса, ондайларды бірнеше қарапайым сұратуларға бөліп қарастыруға болады.

Деректер базасын басқару жүйесінің қосымша функциялары. Әдетте деректер базасын басқару жүйесі бірнеше қосымша қызметтерді қамтиды. Мысалы, ол көбінде

іздестіру негізінде табылған ақпаратты тұтынушыға қолайлы түрде (мысалы, кесте түрінде) – шығарып беруді ұйымдастыра алады. Сонымен бірге деректер базасын басқару жүйесі көптеген есептеу жұмыстарын да орындай алады. Мәселен, мынадай сұратуды: "деректер базасында белгілі шарттарды алдын ала қанағаттандыратын қанша нысан бар?".

Деректер базасын басқару жүйесінде басқа да есептеу жүйелерінің атқаратын кейбір функцияларын өз міндетіне алатын мүмкіндігі болуы қажет. Мысалы, оған жататындар: ақпаратты жүйелік "жаңылысу" (сбой) кезінде пайда болатын қателерден, тұтынушының кездейсоқ қателерінен және қажеттікке қарай кейбір жүйеге рұхсатсыз кіруден сақтандыру жұмыстары.

Деректер базасының теориясы автоматтанған фактографиялық ақпарат іздестіру жүйесін құрудың математикалық және программалық құралдары ретінде көрініс табатынын да айтуға болады.

Дейт К. Введение в системы баз данных. М., 1980.
Информатика / Под ред. К.В.Тараканова. М., 1986.
Ланкастер Ф. Информационно-поисковые системы. М., 1972.
Мартин Дж. Организация баз данных в вычислительных системах. М., 1980.
Московичь В.А. Информационные языки. М., 1971.
Панков И.П., Захаров В.П. Информационно-поисковые системы / В уч. Прикладное языкознание / Отв. ред. А.С.Герд. СПб., 1996.
Солтон Дж. Динамические библиотечно-информационные системы. М., 1979.
Сэлтон Г. Автоматическая обработка, хранение и поиск информации. М., 1973.
Тиари Т., Фрей Дж. Проектирование структур баз данных. М., 1980.
Черный А.И. Введение в теорию информационного поиска. М., 1975.

АВТОМАТТАНҒАН ОҚУ ЖҮЙЕСІНІҢ НЕГІЗГІ ҰСТАНЫМДАРЫ

Оқытуды компьютерлендіру ісі қазіргі кездегі педагогиканың ең бір күшті бетбұрысы деуге болады. Сонымен бірге бұл білім беру жүйесінің алдында тұрған күрделі мәселелерінің біріне айналуда. Тек оқытудың автоматтанған түрі ғана дәлелділік (уәжділік, мотивация) сипатының жоғарғы деңгейінің негізінде оқып-үйрету әсерінің нәтижелігін түбегейлі өзгерте алады. Себебі білім беру барысының ескі (дәстүрлі) әдіс-тәсілдеріне жұмсалатын адам еңбегі (инемен құдық қазғандай) есептеу техникасына, яғни компьютерге жүктеледі. Сондықтан адам баласының бір рет қана атқаратын еңбегі оқып-үйренушінің қалауынша компьютер көмегімен сан рет қайталануына мүмкіндік туады. Әрине, оқытуды автоматтандырудан педагогтың еңбегін мүлде қажетсіз деген ұғым туындамауы керек. Бірақ автоматтандыру әдісі оқытушыны "қара жұмыстан" – шығармашылық әрекеттен босатып, оны қарапайым механикалық амалдарға ауыстырады. Ал мұндай жұмыстарды адам (мұғалім) әрекетімен салыстырғанда ЭЕМ (компьютер) "шаршамай-шалдықпай" зор нәтижелікпен және өте жылдам орындай алады.

Оқытуды автоматтандыру барысы, ең алдымен, **автоматтанған оқу жүйесін** жан-жақты зерттеп, құрастырып алып, тек содан кейін ғана оны қолдануды қажет етеді. Көбінде автоматтанған оқу жүйесі ұғымына оқытуға қажетті техникалық құралдар жиынтығы мен оқыту программаларының дестесін (пакетін) жатқызады. Дестедегі әрбір бағдарлама тек белгілі бір сабақты жүргізуге, яғни оқытушы мен оқушы арасындағы әрекеттесудің бір ғана сеансын өткізуге икемделген. Мұндай тәсіл біржақты сипатқа ие, себебі "оқу жүйесі" ұғымы мынаған саяды: нақты автоматтанған оқыту (үйрету) программалары оқу жоспарына сәйкес келуімен бірге автоматтанған оқу жүйесінің ішкі құрылымдық көрсетімімен тұтастық кешенін құрайтын мағынада түсінік алуы керек. Мұндай тұтастықта қарастыру келесі екі негіздемеге сүйенеді: дидактикалық материалдың ортақтығы және оқытудың стратегиясы мен тактикасын автоматты түрде таңдау. Дидактикалық материал ретінде компьютерлік программаларды пайдаланатын белгілі пәндік саланың немесе кең мағынадағы білім базасының деректері қарастырылады. Екінші негіздік жайт – автоматтанған оқу жүйесі шеңберінде әрекет ету кезіндегі үйренушілердің алдыңғы қызметінде жүзеге асырылған мәліметтері, яғни оқу қызметі нәтижелерінің деректер базасы.

Дей тұра, автоматты түрдегі даралық оқыту программасының да белгілі дәрежедегі құндылық жағын теріске шығаруға болмайды. Бірақ автоматтанған оқу жүйесі құрамына енетін дара оқыту программасының нәтижелігі әр уақытта жоғары сатыда тұратынын ескеру қажет. Сондықтан ондай компьютерлік программалар даралық оқыту мен топтық оқыту процестерінің бір жүйе шеңберінде өзара әрекеттесу мүмкіндіктерін ескере құрастырылғаны жөн болады.

Жалпы жағдайда, автоматтанған оқу жүйесінің құрамы мынадай негізгі компоненттерден тұруы керек (*Жарков, 60)*:

1) оқытудың техникалық құралдар кешені;
2) оқыту программаларының дестесі;
3) пәндік ая бойынша білім базасы;
4) оқыту әсерінің нәтижелерін тіркейтін деректер базасы.

Автоматтанған оқу жүйесі әртүрлі пәндерді оқыту барысында пайдаланылуы мүмкін. Оның қызметі сол жүйедегі программалардың оқушыға тигізетін әсері арқылы анықталады. Бұл программалардың негізгі қызметтеріне жататындар:

- оқытудың өзіне тікелей қатысты жағдайлар;
- жаттығу;
- бақылау;
- бағалау.

Автоматтанған оқу жүйесінің қызметін бұлайша бөліп құрастыру оқыту жобалау мен құрастыру кезінде аса маңызды: программа сценарийінің жалпы көрінісін, оқып-үйренушілердің диалог формасын (егер диалог алдын ала қарастырылса), сонымен бірге белгілі дәрежеде оқып-үйренушінің оқыту программамен өзара әрекеттесу сеансынан бұрынғы және кейінгі әрекеттерін анықтайды.

Автоматтанған оқыту жүйесінің негізгі ұстанымдарының бірі ретінде оның толық сипаты саналатын болғандықтан, яғни аяқталған оқыту үзіндісін жүзеге асыру мүмкіндіктерінде оқыту программаларының үш түрі де *(ақпараттау, жаттықтыру, бақылау)* толық сипатта орын алуы қажет. Тек осы аталған талап түгелімен орындалған жағдайда ғана автоматты түрде оқыту жүйесі толықтық сипатқа ие болады. Бірақ тек бір ғана оқыту программасы шеңберінде әртүрлі қызметтерді қамту алгоритмдерге жүктелетін мәселелердің негізгі айырмашылығына айтарлықтай күмән туғызады.

Жоғарыда аталған үш түрлі программалар үшін дидактикалық материалдар пайдалану тәсілімен айырым тапқанымен, бірдей дәрежеде қамтылуы қажет. Ал бұл жайт дидактикалық материалдардың (программада пайдаланылатын мәліметтердің құрылымы) өзара тәуелсіз болуының бірде-бір маңызды кепілі. Білім базасындағы немесе әмбебап тілде жазылған көрінісіндегі бірыңғай деректер өздерінің түрліше атқаратын қызметіне қарай, программалардың орындалу барысында әрқилы жүзеге асуы мүмкін. Мұндағы "әмбебап тілде жазылған көрініс" дегенді, деректер мен оларды өңдеу әдістерін формалды түрге келтіретін құралдардың жиынтығы деп түсіну қажет. Бірдей деректер қызметінің программадағы айырым табуы адам еңбегін үнемдеу ақиқаттығынан басқа, ол материалдарға (ұқсас және қайшылықсыз) өзгерістерді тек бір рет қана енгізуге мүмкіндік туғызады.

Ақпараттаушы (информирующие) программалар арқылы оқып-үйренушілер (обучаемые) дидактикалық материалдармен таныса алады. Оқып-үйренушіге қажетті ақпарат беру тәсілі деректер жиынын айрықша безендіруді немесе оларды көрнекі түрде экран бетінде орналастыруды, аса маңызды ақпаратты айрықша бөлектеп көрсетуді қажет етеді. Дұрысында, бір қалыптағы "жансыз" мәтіндік ақпаратты қажетті графиктер арқылы, компьютердің әуендік және басқа да құралдары арқылы оларға "жан бітіріп", оқып-үйренушінің жадында сақтау мүмкіндігін арттыру. Оқып-үйренушінің қалауынша бұрын танысқан ақпарат үзіндісіне қайта оралып, жаттығуына да жағдай жасалуы қажет.

Жаттығуға қатысты программалар өз құрылымына бөлек тәсілді қолдануды қажет етеді: оқып-үйренушіні білім базасына немесе оқу материалының ең қиын деген және біркелкі қате жіберілетін үзінділері бойынша түзілген жауаптар жүйесіне айрықша назар аударуға мүмкіндік туғызу. Сонымен бірге оқып-үйренуші мен программа әрекеті арасындағы (және керісінше) диалог жүргізілгені де жөн болады. Мұндай программаның орындалу барысында, оқып-үйренуші тарапынан кеткен қателерді атап көрсету, тіпті программаның күрделігіне қарай, ол қателердің не себепті орын алғанын да хабардар етіп отыру жөн саналады.

Бақылаушы программалардың жұмысы кезінде әрбір тапсырманың жауаптары деректер базасының оқыту барысы нәтижелерінде сақталып отыруы қажет. Сонымен бірге ақиқат және кездейсоқ қателерді аса ұқыптылықпен ажырата білу керек. Бұл жайт лингвистикалық оқып-үйрету программаларында аса маңызды деп саналады. Себебі жаңылыс басылған сөздердің салдары күрделі қателерге әкеп соғуы мүмкін.

Егер программада ***бағалау*** қызметі жүзеге асатын болса, қателерді әрі қарай да талдау қажеттігі туындайды, яғни кеткен қатенің "салмақ" деңгейін анықтау және орындалған тапсырма нәтижелерін статистикалық өңдеуден өткізу қажет болады.

Автоматтанған оқу жүйесінде бірнеше қосымша қызметтерді бөліп алып қарастыруға болады. Ондай қызметтерге төмендегілерді жатқызуға болады (*Жарков, 62*):— оқу қызметі нәтижелерін тіркеу. Егер автоматтанған оқу жүйенің құрамында осындай деректердің базасы орын алса, онда мұндай қызметтің маңыздылығы арта түседі және оқып-үйренушілердің алған нәтижелерін тіркеу жұмысының әсерлігі де барынша өседі.

- оқып-үйренушіні ақпараттау қызметі. Егер білім базасының жүйесі арқылы оқып-үйренуші өз сұранысына қажетті деген ақпаратты алуына мүмкіндік туса, бұл қызмет толық көлемде жүзеге асады.

- диалогты ендіру қызметі. Диалогты пайдалану үшін жан-жақты зерттелген, қажетті деген айтылымдарды құрастыру мен тани білетін табиғи тілге барынша жақын келетін ***жасанды тіл*** құрастыру қажет.

- оқып-үйренуге қажетті стратегия мен тактиканы таңдау қызметі, яғни жүйені оқып-үйренуге икемдеу қызметі. Мұндай икемдеу қызметі үшін оқыту әрекеті бойынша құрылған деректер базасын пайдалану орынды деп саналады.

Аталған қосымша функцияларды оқыту программаларының түріне қарай үлестірілу сипаты да өзгеше болып келеді. Мәселен, нәтижелерді тіркеу қызметін жүзеге асыру тек бақылаушы программаларында ғана мүмкін, ал ақпарттау функциясының мұндай программаларға қатысы жоқ деуге болады.

Оқыту жүйесін құрастыруға байланысты бірнеше қолданбалы мәселелерді қарастырып, осындай қызметтерге қажет деп саналатын бірнеше ұстанымдар (принциптер) төмендегідей (*Жарков, 62-63*):

1) Көптеген автоматтанған оқу жүйесінің жоғары нәтижелігінің жеткіліксіздігі оқып-үйрену ісіне кешенді түрдегі қатынастың (подход) жоқтығы. Оқытудың кешенділік ұғымын ЭЕМ-ның (компьютердің) оқып-үйрену барысына толық араласу жағдайы деп түсіну қажет. Егер автоматты оқытудың компьютерлік программалары үзілді-кесілді сипатта болып, оқу курсының тек бірнеше бөлімдерін ғана қамтитын болса, ол нәтижелі бола бермейді.

2) Тілді оқытып-үйрету (обучение языку) **көпаспектілі** болуы қажет. Осыған байланысты оқыту программасын құрастыру кезінде оқып-үйрету мазмұнын құрайтын ақпаратты тілдің әр деңгейіне бірдей қатысты болуын мақсат етіп қойған жөн болады. Мұндай жағдайда оқыту барысының бірсарындылығы (монотондығы) азаяды. Осының нәтижесінде үйренушіде тек статикалық аспектідегі тіл жүйесі туралы түсінік қана қалыптасып қоймайды, сонымен бірге әртүрлі тілдік бірліктердің қызметі туралы да түсінік пайда болады. Мысалы, оқушы жаңа атаулы (номинативті) бірліктерді меңгеру үстінде оларды бірден әртүрлі синтаксистік құрылымдарда да қолдануына мүмкіндігі бар. Осындай мәліметтермен бірге оқып-үйренуші атаулы бірліктердің валенттілік пен тіркесімдік қасиеттерін де сәйкес игере алады.

3) Оқыту барысының табиғилығы, біріншіден, үйретілетін тілдің ең ірі және жиі кездесетін құбылыстарын көрсете алатын оқу материалдарын мұқият түрде таңдаудан тұрады; екіншіден, үйренуші жағынан басқаруды қажет ететін

программаларда (мысалы, білім базасымен әрекет жасау) компьютермен жұмыс істеуде арнайы дайындықсыз-ақ диалог жүргізуді ұйымдастыру қажет болады; үшіншіден, компьютерлік программамен жүзеге асатын диалог тілі табиғи тілге барынша жақын дәрежеде жүргізілуі керек.

Үшінші жағдай қайсыбір нақтылауды қажет етеді. Көптеген диалог жүргізу мақсатындағы жүйелерде ағылшын тілі немесе соған ұқсас келетін тілдер қолданылады. Бірақ тілдерді оқыту программаларында, әсіресе, жаңадан үйренушілер үшін ана тілінің негіз болғаны жөн. Үйренуге (үйретуге) қажетті тіл ретінде, көбінде, ағылшын тілінің алынатынын ескерсек, оның интерфейсі ретінде, басқаша айтқанда, компьютер мен адам арасындағы диалогты жүзеге асыратын программалық құрал тілі ретінде қолдану үшін ағылшын тілін аздап болса да игеру қажет болады.

4) Табиғилық және вариативтілік принциптері бір-біріне іргелес жатады. Тіл жүйесі, сөйлеу тілінде әртүрлі тілдік құрылымдарды жүзеге асыру арқылы бір ғана немесе бір-біріне ұқсас келетін коммуникативтік міндеттерді көптеген тәсілдер арқылы шешуге мүмкіндік тудырады. Олардың қайсыбірін таңдап алу, оқу барысында бейнеленетін экстралингвистикалық факторларға негізделеді.

5) Оқып-үйренуді автоматтандыру оқу барысында үйренушінің қажеттілік, қызығушылық деңгейін арттыруға мүмкіндік туындатады. Бірақ мұндай уәждік (мотивациялық) жоғары деңгейіне оқу программасының коммуникативтік негізде (яғни машина мен адам арасындағы диалогты жүзеге асыру арқылы) және оның жағдаяттық бағытталуы (бейімделгіштігі, яғни тапсырманың жағдаяттығымен санасу) арқылы ғана жетуге болады.

Автоматтанған оқу жүйесін жобалау мен жете зерттеуде мынадай негізгі міндеттердің шешімін табу қажет болады:

- материалды таңдау және оны топтау (әр сабаққа, тақырыптарға үлестіру);
- диалогтың сценарийін жете зерттеу;
- сыртқы пішін көрінісін безендіру немесе программаның дизайны;
- жүйедегі өзара әрекеттікті ұйымдастыру (оқып-үйретуге кешенді тәсілді жүзеге асыру);
- "деректер көрсетімі тілін" жете зерттеу;
- пәндік сала және онымен әрекет етуге қажетті интерфейс бойынша деректер базасын құрастыру.

Осындай әрбір ішкі міндеттердің әртүрлі оқу программаларының типтеріне қатынасы бірдей емес.

Автоматтанған оқу жүйесінің функционалды өзегін оқыту программалары құрайды. Себебі, үйренушінің оқыту программамен әрекеттесу барысы, шындығында да, автоматтанған оқу жүйесі арқылы жүзеге асады. Міне, осы әрекеттесу ғана оқыту әсерлігінің аса жоғары уәждігі (мотивация) мен нәтижелігін анықтайды.

Оқыту программалары әртүрлі негіздер жиыны бойынша топтастырылуы мүмкін. Олардың ең маңыздылары мыналар (*Жарков, 64-66 бб.*):

1) Оқып-үйренушінің ЭЕМ-мен (компьютермен) тікелей әрекеттесуіне оқыту программасының мүмкіндік тудыруына қарай. Кең мағынада алғанда, оқытуға қатысты программалар түр-түрге бөлінеді, бірақ олардың бәрі бірдей оқыту қызметін атқаруға бейімделмеген. Мәселен, ондай программалар қосымша сипатта мынадай қызметтерді атқара алады: оқу барысы жайлы ақпараттарды статистикалық немесе басқаша өңдеу қызметі, "әкімшіліктің" атқаратын көптеген қызметтерін есептеу құралдарына жүктеу (сабақтардың күнделік кестесін құру, оқып-үйренушілер мен оқытушылардың қызметтерінің нәтижелерін тіркеу және оларды техника жадында

сақтау. Әсіресе, мұғалімдердің көп уақыты мен күшін қажет ететін оқу бағдарламасын жасау әрекеті, есеп беру жұмысын атқару және т.б.

2) Оқыту процесінің аяқталған көрінісін жүзеге асыру үшін компьютерлік программа қандай қызметтерді атқаруына қарай: бақылау, қажетті әрекет түрін көрсету, оқып-үйренушінің әрекетін бағалау және т.б. Аталған міндеттердің әрбіреуін шешетін алгоритм тек оның өзіне ғана тән ерекшелікке ие, сондықтан олар бір ғана алгоритмдік тәртіп бойынша жүзеге аспайды; әртүрлі программалар үшін ортақтық сипат, ол оқыту барысына ешбір қатысы болмайтын кейбір таза формалды әрекеттер (процедуралар). Мысалы, терезелік интерфейс, жинақтауыштағы деректермен алмасу процедуралары, қыстырма (заставка) және басқа да программаның безендіру құралдары, оқып-үйренушімен диалог ұйымдастыру құралдары және т.б.;

3) Ұсынылатын оқу міндетінің типіне қарай: дағдылы (обычное), тапсырмаға ену (погружение); тапсырманың жағдаятына ену;

4) Оқу программасында диалогтың орындалуына және оның формасына қарай. Шындығында, автоматтанған оқу жүйесінің уәждік сипатын күшейтетін жеткілікті түрдегі маңызды құрал ретінде диалогты алуға болады; бірақ ол тағы да бір аса маңызды қызметке ие: бұл программа – оқып-үйренушінің "кері байланысын" (обратная связь) жүзеге асыратын бірден-бір құрал. Осының негізінде қандай да болсын формадағы күрделі оқу программалары диалогсыз жүзеге аспайды. Оқып-үйренуші жағынан да программаның орындалу барысын басқару және оның даралық ерекшелігін де есепке алу жұмыстары диалогсыз мүмкін емес. Диалог ескерілмеген оқыту программаларымен жұмыс жүргізу кезінде оқып-үйренуші әбден енжарлық күйде қалуға мәжбүр болады.

Диалог қайсыбір табиғи тілдің (ана тілі немесе ағылшын тілі) арнайы құрастырылған бөлігіне қатысты немесе жасанды бұйрықтар жүйесі негізінде құрылуы мүмкін. Анығы, оқыту программаларында табиғи тілді пайдалану қай жағынан алғанда да аса құнды, бірақ жасанды интеллект деңгейіне көтерілу қажет болатындықтан, ол көптеген қиындықтарға әкеп соғады; сонымен бірге оқып-үйренушілер бұйрықтардың арнаулы тілін алдын ала біліп алуы қажет болады және мұндай тіл тек бірен-саран оқыту программаларында ғана қолданылып қоймай, оның бүкіл оқыту жүйесінде орын алғаны аса мәнді.

5) Оқыту программасының орындалу барысын оқып-үйренушінің өзі басқара алатын мүмкіндігіне қарай. Бұл жерде сызықтық (линейное) және тармақты алгоритмдер туралы сөз болып отыр. Бірінші жағдайда, яғни сызықтық алгоритмдерде программаның сценарийі бойынша оқып-үйренушілердің алдын ала берілген бақыланылатын және ұқсас әсерлер реттілігінен ауытқуға ешбір рұхсат жоқ. Екінші жағдайда, яғни тармақты алгоритмдер жағдайында аталған реттілік оқып-үйренушінің қалауынша құбылып тұруына мүмкіндігі бар. Бұл жағдай оқып-үйренушінің оқу барысына қатысының белсенділігін арттырады.

6) Оқып-үйренушінің даралық ерекшелігінің ескерілуіне қарай. Даралық ерекшелік дегенде оқушының өзіне ғана тән сапалық сипатынан гөрі, оқып-үйренушінің деңгейіне қатысты қайсыбір сандық сипаттама деп ұғынған жөн. Өзара айырым табатын деңгейлер саны барынша үлкен болуы да мүмкін (кейбір жүзеге асырылған автоматтанған оқу жүйесінде оның саны тоғызға дейін жетеді). Даралық ерекшелікті ескеру оқыту тәсілінің дербес болуын таңдаудан туындайды, яғни ол толығынан алғанда, оқу барысының нәтижелігін барынша ұлғайтуға мүмкіндік жасайды. Мұндай әр оқып-үйренушінің дара мүмкіндігін есепке алу, программаның іске қосылмай тұрған кезінде, яғни алдын ала және программаның

орындалу барысында да ескеріледі. Соңғы жағдайда бір деңгейден екінші деңгейге өтуге мүмкіндік туады. Мұндай ауысу әрекеті аса қарапайым немесе аса күрделі тапсырманы таңдайтын оқып-үйренушінің қалауы бойынша да жүзеге асуы мүмкін және сонымен бірге программаның орындалу барысында алдыңғы әрекеттердің ойдағыдай жүруін автоматты талдау арқылы да жүзеге асыруға болады. Деңгейді бағалауда субъективтілікті (өзіндік сипатты) жою тұрғысынан автоматты талдау тәсілін қолдану ең ұтымды деп саналады;

7) Оқып-үйренушінің білім базасымен әрекет ету мүмкіндігіне қарай. Әрине, мұндай мүмкіндік білім базасымен әрекет жасауды жүзеге асыратын программаларды ғана сипаттай алады, яғни ақпараттаушы, оқып-үйретуші және жаттықтырушы программалар бақылау жұмысына жарамсыз болып саналады.

8) Автоматтанған оқу жүйесінде программалардың орындалу нәтижелерінің тіркелу жағдайына қарай. Бұл мүмкіндік өзінің басқа да құндылықтарымен бірге нақты оқып-үйренуші мен автоматтанған оқу жүйесі арасында өтетін сеанстық жұмыстар әсерінің сабақтастығын қамтамасыз ете алады. Мысалы, әрбір оқып-үйренушіге қатысты қайсыбір программаның орындалу жағдаяты тіркеледі. Ал бұл оқыту барысында кездесіп тұратын кейбір тақырыптың сырт қалуынан және қайталануынан сақтандырады. Сонымен бірге нәтижелерді тіркеу ісі оқып-үйренушінің білім деңгейін оқу барысында ескеріп, реттеп отыруға мүмкіндік тудырады.

9) Дидактикалық материалдардың программа мәтінінде немесе оған тәуелсіз деректер берілісінде орын алуына қарай. Мұндай жағдайда әртүрлі деректер жиынтығымен бір программа арқылы әрекет етуге мүмкіндік туады. Осының нәтижесінде оқыту жүйесін программалауға жұмсалатын адам еңбегі мен оқу-үйрену материалдарына тиісті өңдеулер жүргізу уақыты үнемделеді.

Осы айтылғандарға байланысты автоматты оқу жүйесінде көптеген міндеттерді шеше алатын және икемділігі мол деректер көрсетімі тілін (язык представления данных) зерттеп құрастырудың маңыздылығы арта түсуде.

Автоматизированные обучающие системы на базе ЭВМ / Под. ред. А.Ф.Чернявского. Минск, 1980.

Жарков И.В. Автоматизированные обучающие системы // Прикладное языкознание: Учебник / отв. ред. А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. С. 59-67.

Носенко Э.Л. ЭВМ в обучении иностранным языкам в вузе. М., 1988. С. 103.

Пиотровский Р.Г. Компьютеризация преподавания языка. Л., 1988.

Рыбаков Ф.И., Соколова И.Н., Лоскутникова Л.В. Программированное обучение с использованием ЭВМ. М., 1968.

Савельев А.Я., Новиков В.А., Лобанов Ю.И. Подготовка информации для автоматизированных обучающих систем. М., 1986. С. 174.

Фадеев С.В. ЭВМ в преподавании русского языка как иностранного. М., 1990. С. 81.

АВТОМАТТЫ (МАШИНАЛЫҚ) АУДАРМА

Сөйлеу әрекетінің қайсыбір түрі қайта туындалатын автоматты жүйелер кибернетикалық модельдер тобына жатады.

Жалпылама айтқанда, адам баласының сөйлеу әрекетінің негізгі үш типін бөліп қарастыруға болады. Ал оларды түпнұсқаға барабар туындату үшін кибернетикалық модельдердің де ең кем дегенде үш типі құрастырылуы қажет:

1) коммуникация кезіндегі адамдардың тілдік әрекеті. Бұл жердегі коммуникация процесін кең түрде ұғынуға болады, яғни адамдардың бір-бірімен тікелей тілдесуін ғана емес, сонымен бірге белгілі ортада – баспа, радио, теледидар және т.б. арқылы тілдесуін де қамтиды. Осы типке қатысты тілдік әрекеттің негізгі формасы – әртүрлі эксперттік жүйелерде модельденетін диалог түрі;

2) қайсыбір білім саласын оқыту жағдайындағы, тілге үйретуді қоса есептегендегі тілдік әрекет. Мұндай әрекет типі оқыту жүйелері арқылы модельденеді;

3) шет тіліндегі мәтінді ана тіліне және керісінше аудару тұсындағы адамның тілдік әрекеті.

Машиналық аударма жүйелері кибернетикалық модель жасау аясына қатысты болуы себепті тілдік әрекеттің үшінші типі арқылы іске асады. Сондықтан аударма ісі электронды-есептеу машиналарының (ЭЕМ) көмегімен аудармашы әрекетін қайталауға бағытталады.

Аудармашының тілдік әрекетінің, яғни сөйлеу-ойлау қызметінің моделін машиналық аударма жүйесі арқылы жүзеге асыру екі түрлі таптастыру параметрлері бойынша анықталады:

- автоматтандыру дәрежесі, яғни компьютерге жүктелетін жұмыстың нақты көлемі. Мұнда мынадай жағдаяттарды айырып қарастыру керек (*Кулагина, 1979, 3)*:

1) аударма жұмысы толығымен немесе толыққа жуық ЭЕМ арқылы орындалады;

2) аударма жұмысын аудармашы атқарады да, ал ЭЕМ жәрдемші рөлін атқарады (мысалы, қажетті сөздің аудармасын сөздіктен тауып алу жұмысында).

- аудармашы адамды қатыстыру стратегиясы, яғни аударма жұмысының негізгі бөлігі компьютер арқылы орындалады да, ал аудармашы аударылған мәтінді өңдеуге, яғни аудару кезінің әр тұстарында (басында, ортасында, аяғында) редакциялық жұмысқа қатынасады (*Machine Translation, 1987, 12-13*).

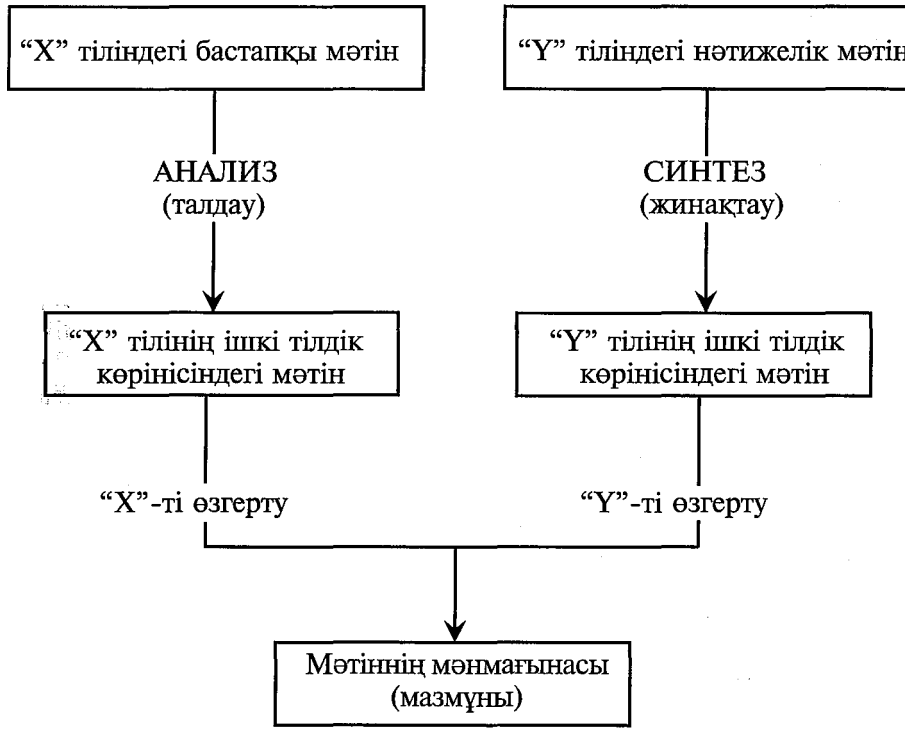
Аудармашының сөйлеу әрекетінің ерекшелігі мынаған саяды: бастапқы нысан ретінде бір тілдегі мәтін алынады да, ал тілдік қызметтің нәтижесі ретінде – мағыналық жағынан аударылуға тиісті мәтінге барабар, басқа тілдегі мәтін болып табылады. Басқаша айтқанда, аталған нәтиже міндетті түрде мәтін түрінде ұсынылуы қажет. Ал басқа бір тілдік әрекет типтерінде, мысалы, бұлшық ет күші арқылы немесе ойлау қызметі, білімді қорландыру және т.б. әрекеттерде нәтиженің мәтін түрінде болуы міндетті емес.

Аталған ерекшелік **автоматты аудармаға** белгілі принциптік құрылымды сақтауды міндеттейді. Төменде ондай құрылым мүмкіндігінің ең жақсы деген (идеалды) көрінісі берілді.

"Х" тіліне талдау жүргізу мен оны өзгертудің мақсаты – "Х" тілінің сөздігі, грамматикасы және қолдану ережелері негізінде "Х" тіліндегі мәтіннің мазмұнын ашу. Ал "Y" тіліне жинақтау әдісін қолдануды – "Y" тіліндегі шығару мәтінін

құрастыру (немесе оны синтездеу) деп атайды. "Y" тіліндегі мән-мағына сол тілдің сөздігі мен грамматикасы негізінде ашылады.

ЭЕМ-сыз дәстүрлі аудармашының мәтінді аудару кезіндегі талдау мен жинақтау компоненттерінің орын алуы айқын түрде байқалмағанымен, **автоматты аударма** жасау процесінде ондай тәсіл бөліктері міндетті түрде орын алуы қажет.



Сызба. Автоматты аударма процесінің идеалдық құрылымы

Ал сызбадағы "өзгерту" топтамасының қажеттілігі ең қарапайым жағдайда да айқын байқалады. Мысалы, аударуға қажетті "X" тілінің сөздік бірлігі мағыналық жағынан екінші тілдің ("Y") сөздік бірлігімен және екі тілдің ("X" пен "Y") синтаксистік құрылымдары да бір-бірімен сәйкес келіп тұрған жағдайда да (тіпті сөздердің орналасу тәртібіне дейін) тиісті өзгерістердің жасалатыны белгілі. Негізінен алғанда, сөздердің грамматикалық тұлғалары мен аударма баламасының сөйлем ішінде дұрыс орналасу тәртібі өзгертуді қажет етеді.

Бірақ жоғарыда қарастырылған жайт идеалды жағдаятына сай және тілдің шектеулі сөздік бөлігі мен синтаксисіне тән болып келеді. Сонымен бірге мұндай жағдаят тек әрбір нақты тілдер жұбы үшін ғана ұтымды деуге болады. Әзірше екі тілдің лексикалық бірліктері мен синтаксистік құрылымдары бойынша басым көпшілік тілдерге жоғарыда аталғандай тұспа-тұс сәйкестік жайлы айту қиын. Оған себеп – әртүрлі тілдердегі көп мағыналы сөздер көлемінің сәйкес келмеуі. Лексикалық сәйкестіктің ең соңғы бұзылу жағдайы – қайсыбір тілдерде баламасыз лексиканың болуы. Әдетте, олар тұрмыстық бұйымдарды, мәдениетті сипаттайтын сөздердің екінші (аударуға қажетті) тілдің лексикалық қорынан орын алмайды.

Мұндай жағдай лексикалық элементті сипаттау немесе калькілеу үшін оның мәнін дұрыс түсінуді қажет етеді. Басқаша айтқанда, мағынаға сәйкес келтіру үшін қажетті деген өзгертулерді (өңдеулерді) жүргізу керек болады.

Егерде аударма мәселесін, *автоматты аударманы* қоса есептегенде, тіларалық сәйкестіктің бұзылуы тұрғысынан қарастырсақ, онда оның мақсатын екі тіл жұбы мәтіндерін мағыналық сәйкестікке келтіру деп ұғыну қажет. Яғни екі тіл арасындағы әртүрлі элементтердің (сөздік, құрылым, сөздер тәртібі және т.б.) мазмұнды ашу құралдарын сәйкестікке келтіру деп түсіну керек болады. Мұндай сәйкестікке келтіру мәселесін шешу жоғарыда көрсетілген сызбадағы топтамалар қызметінің барлық машиналық аударма жүйесінде орындалуын қажет етеді. Сәйкестікке келтіру мүмкіндігі және аударманың принциптік жақтарының барлық тілдерге ортақ болуы тілдің сызықтық және таңбалық сипаттарына негізделеді (Ф. де Соссюр). Ал таңбалық қасиеттің негізінде (мағыналық тұрғыда) барлық халықтардың ойлау қызметін таңбалау мен қоршаған ортаның физикалық ортақ қасиеттерінің жататыны белгілі.

Машиналық аударманың өмірге келуіне, біріншіден, XX ғасырдың екінші жартысынан бастап әр елдерде (континенттерде) бірнеше тілдегі ақпарат ағымының қарқындап өсуі, екіншіден, ғылыми-техникалық прогресс үшін оларды меңгеру қажеттігіне қатысты әлеуметтік себептер негіз болды. Осыған байланысты машиналық (автоматты) аударма жұмысының даму барысына қатысты кейбір жайттарға тоқталайық.

Аталған әлеуметтік себептер аударма мәселесінің шешімін табуды және оның ұтымды жолдарын іздестіруді қажет етті. Бұған ең алдымен аудармашының тілдік қызметінің ерекшеліктері ішінен шеттілдік мәтіннің құрылымын зерттеудегі таза формалды талдау жолын жатқызуға болады (*Панов, 1958).* Мысалы, аудармашыға мәтіннің грамматикасы және лексиканың көмекші сөздері мен есімдік сөздері және басқа да осы сияқты функционалды грамматиканың элементтері белгілі болғанымен, салалық мәтіннің арнаулы лексикасы таныс болмауы мүмкін.

Шеттілдік мәтінді аударудағы оның тұрпаты (формасы) бойынша мән-мағынаны ашу ерекшеліктерін модельдеу, ең алдымен, аудармашының іс-әрекетін зерттеу нысаны ретінде алуға бірден-бір себепші болды. Сонымен, мұндай модельдеу шеттілдік мәтінді аударуда тілдің ақиқат деректерін "шеттетіп", тілдің семантикасын арнайы талдамай-ақ, формалды тұрғыда аударма баламасын табуға мүмкіндік жасайды.

Автоматты аударма жасаудың бастамасы ретінде ресми түрде 1949 жыл саналады. Тілді формалды тұрғыда қарастыра отырып, электронды-есептеу техникасының көмегімен аударманы жүзеге асырудан бұрын оны механикалық машиналар жолымен іске асыру идеясы аталған мезгілден де көп бұрын бастама алғаны белгілі. Мысалы, А.Кахердің ойлап тапқан аударма-жазба машинасы жайлы мәлімет "Yава таа" газетінде 1924 жылдың 24 ақпанында шағын хабарлама ретінде жарық көрген болатын (*Переводная..., 1962).* 1933 жылы Кеңес Одағында П.П.Смирнов-Троянскийге бір тілден екінші тілге аудару машинасын ойлап шығарғаны үшін арнайы патент тапсырылғаны да жұртшылыққа мәлім (*Переводная..., 1959).*

Аударма жасауда екі немесе одан да көп тілдердің бірліктері арасындағы мағына сәйкестігін анықтау үшін алғашында механикалық тәсілдер қолданылды. Сондықтан аударма жұмыстарының жетістіктері әлемдік ғылымға тек XX ғ. 50-жылдарында ғана белгілі бола бастады. Ең алғашқы машиналық аударма жүйелері АҚШ-та пайда болды. Тарихи тұрғыдан қарастырғанда машиналық аударма жүйесін құрастыруға жалпытілдік көзқарастағы криптография (шартты белгілермен жазылған құпия жазу) саласының теориясы мен практикасы негіз болды.

Екінші дүниежүзілік соғыс кезінде әр елдерде, әсіресе Англия мен АҚШ-та құпия хаттардағы шифрлы жазуды үйреншікті жазуға ауыстыру (дешифровка) жұмысын машинасыз да және машинамен де әрекет ету арқылы көптеген нәтижелерге қол жеткені белгілі. Мысалы, гитлерлік басқару орындарының шифрланған бұйрықтары мен хабарларын жай жазуға айналдыратын "Ультра" атты дешифровалды машинаның құрастырылғанын айта кетуге болады.

Лингвистикалық тұрғыдан қарастырғанда, шифрланған (кодталған) хабарды дәстүрлі жазуға келтіру дегеніміз, ол лексикалық қоры белгісіз және грамматикалық ережелері әлі анықталмаған шет тілінде жазылған мәтінді аудару жағдаятымен барабар деуге болады. Мұндай аударма жұмысы бір кодтағы жүйеден екінші кодтағы жүйеге ауыстыру әрекетімен бірдей келеді.

Маман-дешифровщиктердің пайымдауынша, шифрланған (кодталған) хабарды жай жазуға келтіру әрекетіне қарағанда бір табиғи тілдегі мәтінді екінші табиғи тілге аудару әрекеті жеңіл деп бағаланады. Себебі, қазіргі табиғи тілдерді сипаттайтын көптеген сөздіктер мен грамматикалардың негізінде тілдің өзіне ғана тән лексикалық, морфологиялық және синтаксистік ерекшеліктері алдын ала белгілі деуге болады. Ал есептеу техникасы ондай сөздіктер мен грамматикаларды қалаған тәртіпте, бағдарламаға сай өз жадында орналастыра алады.

1946–1947 жылдары ЭЕМ арқылы машиналық аударма жасау идеясы Принстонның ғылыми-зерттеу институтында арнайы мәселе ретінде талқыланады (У.Уивер, А.Л.Бут және т.б.). Ең алғаш бұл идея дешифровка ісінің американдық маманы Уоррен Уивердің меморандумында бірізді шешімге келтірілді. Ал 1949 жылдың 15 шілдесінде атақты ғалым өз меморандумында көтерілген мәселелерді тіл және есептеу техникасы мен дешифровка салаларындағы екі жүзден астам мамандарға жария етті.

У.Уивер машиналық аударманың принциптік мүмкіндігін мүлде бөлек тілдердің ұқсас қасиеттерімен және әр тілдің өзіне ғана тән формалды сипатта болуымен байланыстырады. Қорыта айтқанда, У.Уивердің пікірі бойынша, қытайша жазылған кітапты қытайша шифрланған (кодталған) ағылшынша немесе басқа ұлттық тілдердегі кітап деп түсінген жөн (*Машинный перевод, 1957, 33-45).*

Сонымен машиналық аударманың өзінше бөлек ғылыми бағыт ретіндегі даму кезеңі Уивердің меморандумынан бастама алады. Бұл бағытты, шартты түрде, негізгі үш кезеңге бөліп қарастыруға болады (*Hutchins, 1986; Кулагина, 1979; Марчук, 1983, 91; Апресян және т.б., 1989, 4-5*). Мысалы, ол хронологиялық жағынан былайша айырым табады:

1-кезең: 1949 жылдан бастап 60-жылдардың ортасына дейін;

2-кезең: 60-жылдардың ортасынан 70-жылдардың аяғына дейін;

3-кезең: 80-жылдардан қазіргі кезеңге дейін.

Мұндай үш кезеңге бөлуге негіз болған ғылым мен техниканың дамуындағы қол жеткен нәтижелер екені белгілі. Міне, солардың ішінен лингвистикалық, математикалық, техникалық машиналық аударма жүйесін жабдықтау мәселелері айрықша орын алады.

У.Уивердің меморандумымен жарияланған бірінші кезең 1949 жылы Вашингтон университетінде құрылған ғылыми орталықтың зерттеулерінен (ағылшынша-орысша машиналық аударма жүйесі) бастама алады. 1950 жылы А.Д.Буттың ғылыми жетекшілік етуімен французша-ағылшынша аударма орталығы құрылады. Кеңес Одағында 1955 жылы Мәскеудегі КСРО ҒА-ның В.А.Стеклов атындағы Математикалық институтында О.С.Кулагинаның жетекшілігімен ең бірінші машиналық аударма орталығы ашылады.

Осы кезеңде ең алғашқы зерттеушілерді үміттендіретіндей нәтижелер алынып, алғашқы машиналық тәжірибелер жүргізілген болатын. Мысалы, 1954 жылы АҚШ-та Джорджтаун университетінде проф. Леон Е.Достердің жетекшілігімен GAT жүйесі негізінде құрастырылған орыс тілінен ағылшын тіліне аударатын машиналық аударма жүйесі жұртшылық назарына ұсынылды. Бұл бір тілден екінші тілге ЭЕМ (IBM-701) арқылы орындалған ең алғашқы аударма тәжірибесі болып саналады. Ғылымда бұл тәжірибе **Джорджтауындық эксперимент** деген атпен тарихта қалды (*Машинный перевод, 1957, 171*-б.). Аталған экспериментте "тікелей" машиналық аударма жағдаяты (бір тілдік жұпқа қатысты жүйелер) тексерістен өткен болатын. Жүйе сөздігінің көлемі бар болғаны 250 сөзді ғана қамтыса, ал оның грамматикасы тек алты синтаксистік ережеге ғана негізделеді (*Machine Translation, 1987, p. 22-23).* Кеңес Одағында француз тілінен орысшаға аударатын машиналық аударманың бірінші эксперименттік жүйесі О.С.Кулагина мен И.А.Мельчуктың авторлығымен 1955–1956 жылдары құрылғаны осы сала мамандарына белгілі (*Кулагина, Мельчук, 1956).* Бұл жүйедегі негіз сөздер сөздігі 1236 сөз, ал сөз айналым сөздігі 250 бірлікті қамтиды. Аударма бірлігі ретінде толық сөйлем түрі есептеледі. Морфологиялық талдау алгоритмі жалғаулардың кестесімен әрекет ету арқылы аударылатын фразаның (сөйлемнің) сөздеріне қажетті деген грамматикалық ақпаратты тіркейді. Бұл жүйенің артықшылығы – омонимдік сөздерді ажырата алатын топтамасы (блогі) болуында. Орыс тіліндегі фразаны синтездеу топтамасында сөздердің орналасу тәртібінің дұрыстығын тексеру мүмкіндігі алдын ала қарастырылған. Сонымен бірге, Т.Н.Молошнаяның жетекшілігімен ағылшынша-орысша аударма жасау жүйесінің алғашқы жобасы ұсынылды (*Молошная, 1960).*

Жаңа бағыттың қарқынды дамуы көптеген мәртебелі ғылыми конференцияларда, симпозиумдарда, жарияланымдарда көрініс тапты және машиналық аударма проблемасымен айналысатын бірнеше ірі орталықтар, топтар мен зертханалар ашылуына себепші болды. Өткен ғасырдың 60-жылдарының басында олардың әр елдегі сандық көрсеткіші мынадай: ФРГ – 22, АҚШ – 21, КСРО – 10, КНР мен Жапония – 6-дан, Франция – 5, Чехословакия – 4, Ұлыбритания мен Венгрия – 3-тен, ГДР мен Италия – 2-ден. Осындай топтар Бельгияда, Болгарияда, Израильде, Канадада, Люксембургте және т.б. елдерде де пайда бола бастады (*Автоматический перевод, 1967, 509).*

Машиналық аударма жұмысының бастамасы жалпы алғанда өрлеу сипатында болды – тілшілер зерттеудің жаңа бағытын ашты, ал программистер күн тәртібіндегі күрделі және аса қызықты мәселелердің шешімін табуға қызыға кірісті. Бұл кезеңде тіл мен аударма процесіне қатысты барлық мәліметтерге қол жеткендей сезінілді. Сондықтан егер оларды формалды пішінге келтіру ісі жүзеге асырылса, машиналық аударманың әмбебаптық жүйесі құрылып, алға қойылған негізгі мақсатқа тез арада жететіндей көрінген болатын. Бірақ бастапқы және кейінгі тәжірибелердің нәтижелері бұл мәселенің шешімі ойлағандай оңай табылмайтынын байқатты.

Солай бола тұра, атқарылған жұмыстар тек машиналық аударма аясындағы зерттеулерге ғана емес, сонымен бірге ғылыми ізденушілерді тілдің теориялық және құрылымдық жағына да көңіл аударуға ынталандырды.

Қорыта айтқанда, автоматты аударманың бірінші кезеңі белгілі дәрежеде айтарлықтай нәтижелі болды. Себебі, аталған кезеңде көптеген іргелі мәселелердің бастамасы анықталып, зерттеудің болашағы сараланып, жолдары айқындалды. Мысалы, автоматты сөздіктерді құрастыру, **аралық-тілді** құрастыру, сөздердің омонимдік сипаты мен синтаксистік құрылымы, яғни сөздердің сөйлемдегі тіркесімі, фразеологизмдердің жасалу жолдары және көптеген басқа да мәселелердің сыр-сипаттары анықталды.

Осы мезгілдерде, бүгінгі таңда да аса маңызды деп саналатын, синтаксистік талдаудың алгоритмінің алғашқы нұсқалары зерттелді: синтаксистік тәуелділік пен тікелей құрастырушыларды, болжамдық талдауды, көпвариантты синтаксистік талдауды, "фулькрумдер" әдісімен талдауды, конфигурациялық талдауды және т.б. әдіс-тәсілдерді ескере отырып, талдау жүргізу жақтары қарастырылды.

Терминтану саласындағы бағыныңқылықтың моделі мен тікелей құрастырушылардың моделін талдаудан толық сөйлемнің құрылымдық байласымын анықтауға болатындығы байқалды. Мұндай құрылымның көпмағыналығының басын ашып алу үшін әртүрлі "сүзбелердің" әдістері қолданылады (қара: ЛМУ-дың математикалық лингвистика зертханасындағы С.Я.Фитиалов пен Г.С.Цейтиннің жетекшілік етулерімен жүргізілген жұмыстар; GETA атты Гренобльдық топ – Centre d'ïtudes pour la tradution autmatigue, жетекшісі В.Вокуа). "Күшті" сүзбе ретінде Д.Е.Хейстің байланыс "күші" ұғымы ұсынылды (The Rand Corporation Automatic Language Data Processing Croup. Santa Monica (Calif.)). Байланыс күшінің жоғары дәрежесіне қарай талдау варианттарының артықтық деңгейі айырым тапты (*Автомат. перевод, 1971. 41-б.).*

Болжамдық талдау сөздік мақаланың айрықша құрылымына негізделеді (Национальное бюро стандартов США, И.Родес; Вычислительная лаборатория Гарвардского университета (Масс.), рук. А.Эттингер). Мұнда әрбір сөзге оның барлық байланыс мүмкіндігі жайлы ақпарат тіркеледі. Талдау тәртібі (процедурасы) мәтінді солдан оңға қарай тексеріп шығу тәсілі бойынша жүргізіледі. Осының нәтижесінде әрбір кезекті сөздің "өлшемінің" алдыңғы сөз бойынша жасалған болжамға сай келетіндігі тексеріледі. Егер ол жасалған болжамға сай келмесе, ондай сәйкестікті іздестіру сол жаққа қарай ауысады. И.Родестің ғылыми тобында бірвариантты талдау қарастырылады, яғни талдауға түскен сөздің сол жағында орналасқан бірінші сөз сәйкестік "иесі" деп саналып, кез келген сөйлемге тек бір ғана синтаксистік құрылым сәйкес қойылады.

Сөйлемнің синтаксистік құрылымының бірнеше варианттарын есепке алуға мүмкіндік туғызатын көпвариантты талдауды бір қадам алға басу деп санауға болады (С.Куно, А.Эттингер, Гарвард университетінің Есептеу зертханасы). Көпвариантты талдауда сөздің синтаксистік потенциалды мүмкіндіктерінің молдығының арқасында синтаксистік байланыста тұратын сөздерге болжам жасауға болатындығы негізге алынады. Көпвариантты талдаудың алғашқы грамматикасы 2100 ереже мен 82 синтаксистік болжамды қамтыды. Сөйлем құрылымының диаграммасы тікелей құрастырушылар моделі негізінде құрылды. Осының нәтижесінде мүмкін болатын барлық варианттар іріктеліп, олардың көпшілігі әртүрлі сүзбелер көмегімен сан жағынан шағындалды. Болжамдық талдау алгоритмінде көпвариантты фразаның синтаксистік құрылымына тән екімәнділік ақиқаттығынан гөрі, оны таңдалып алынған варианттарды іріктеу әрекетінің формалды сипатының салдары деуге болады. Міне, бұл жүйелердің осалдығы да осында. Дегенмен, синтаксистік болжам жайлы идеяның өзі де құнды деуге болады. Бұл ой-пікірге қазіргі кездегі кейбір грамматикалық болжамдарға сүйенетін автоматты талдау жүйелері негізделуде (*Цейтин және т.б., 1982; Откупщикова, 1982).*

Атап айтуға тұратын жайт – бірвариантты пен көпвариантты талдауларды бір-біріне қарсы қою машиналық аударманың қалыптасуының бірінші кезеңі үшін ерекше маңызды болуы. Бірвариантты талдауда кездейсоқтыққа жақын дәрежеде таңдалып алынған сөйлем құрылымы дұрыс болмай шығуы мүмкін, ал талдау жұмысын басынан қайталау мүмкіндігі бұл жүйелерде алдын ала қарастырылмаған. Көпвариантты талдауда біркелкіліктің бұзылу қаупі бар: құрастырылған

құрылымдардың саны вариантты дұрыс таңдау мүмкіндігінен де айтарлықтай үлкен болуы мүмкін.

Сонымен бірге жақсы аудармашы (немесе автоматты аударма жүйесі) мәтін мазмұнын түсіну үшін аударуға тиісті фразаның дұрыс түрдегі бір ғана вариантын ұсынуы керек. Ал фразалардың көпмағыналығы мәнмәтін бойынша өз шешімін табатыны белгілі, сондықтан сөйлемдердің нақты мағыналарын ажырату мәселесі мәтінде баяндалатын ақиқат жағдаятқа қатысты толық мәтіннің синтаксистік және семантикалық заңдылықтарына тікелей байланысты болып келеді. Мұндай семантикалық мәліметтер машиналық аударманың бірінші және екінші кезеңдерінде үзінділер түрінде ғана көрініс тауып, толығымен қолданыла қоймады. Себебі, машиналық аударма жүйесін тиісті лингвистикалық мәліметтермен жабдықтау ісі әлі де болса толық зерттелмеген болатын.

Конфигурациялық талдау әдісі аударылуға тиісті мәтінді (жүйеге енетін мәтін) сол тілдің алдын ала іріктеліп алынған синтаксистік конструкциялар жиынымен салыстыру тәсіліне негізделеді (*Каф. матем. Лингв. ЛГУ; Белопольская, Ковригин, 1963; Воронин, 1961. Каф. Струк. и прикл. Лингв. МГУ; Дрейзин, Рашитов, 1961. Группа авт. перевода ИЯ АН СССР; Молошная, 1960 және т.б.).* Егер аударма жүйесіне енетін мәтіннің анықталған конфигурациясы жиынның синтаксистік конфигурацияларының біреуімен сәйкес келсе, онда мәтіндік конфигурация танылды деп саналады және ол әрі қарайғы талдау әрекетінде ықшамдалған түрінде көрініс табады. Синтездеу әрекетінде анализдеу кезінде анықталған конфигурация аударуға тиісті тілдің оған сәйкес келетін конфигурациясымен салыстырылады.

"Фулькрумдер" әдісі бойынша жүргізілетін талдауда мәтіндегі сөйлем төменгі сатыдағы бөліктерге ажыратылады. Осының негізінде кез келген фразаны циклдық әдіс бойынша бірнеше рет қарастыра отырып, іріленген фрагменттерге біріктіруге болады. Әрбір кезекті циклдық қарастыру кезінде талдаудың тірек нүктелері (fulcra) – фулькрумдер анықталады. Анықталған фулькрумдер синтаксистік конструкциялардағы негізгі және "басқарушы" сөздер тобына жатады. Олардың құрылымы мен шекарасын анықтау циклдық қайталаулардың мақсатына енеді (*группа автомат. перев. фирмы Bunker Ramo-Woldridge Inc, Canoga Park, Calif., рук. П.Гарвин).*

Тізбекті талдау әдісі арқылы фразаның синтаксистік құрылымын тану үшін берілген сөйлем көрінісіндегі синтаксистік категориялар тізбегін элементарлық тізбектерге бөлу қажет болады (*Харрис, Н.Сейгер*).

Корреляциялық талдау кезінде әрбір сөзге индекстер жиынтығы тіркеліп жазылады. Олардың әрбіреуі сол сөздің осыған ұқсас индексі бар басқа сөздермен өзара байланыстылық (корреляция) мүмкіндігін бейнелейді. Корреляцияның (өзара қатынастылық) жоғары деңгейі лингвистикалық шектеулер арқылы төмендетіледі. Мысалы, сөз, семантика және т.б. сатыларға төмендейді.

Осымен бірге зерттеушілер машиналық аударма мәселесі үшін толық мәтінді талдау мәселесін қарастыра бастады. Соның ішінен толық мәтіндердің антецедент – зат есімдермен арақатынасын анықтайтын ережелердің қарастырыла бастағанын атап кетуге болады (антецедент – анафоралық қатынас. Мәселен, есімдіктен бұрын тұратын есім сөз, ал есімдік осы есімге нұсқайды) (*Шумилина, 1961; Автомат. перевод, 1967, с. 235 к сл.).*

Қазіргі кездегі жоспарланып жатқан машиналық аударма жүйесі екі топқа бөлінеді: бір тілден екінші тілге аударатын – **бинарлы** және бірнеше тілдерден бір тілге немесе бірнеше тілдерге аударуға арналған – **көптік**. Машиналық аударма жүйелерінің бірінші түрінің әр уақытта артықтық сипаты адам баласының аудару

әрекетін модельдеудің және әрбір тілдер жұбы деректерінің бинарлық сәйкестіктерін анықтаудың теориялық және практикалық мүмкіндіктерінің табиғилық жағдаятынан туындайды. Машиналық аударманың бинарлық жүйелерінің кемшілігі – аударуға тиісті "кіретін" тілге талдау жүргізу кезінде олардың қасиеттеріне қатаң түрде тәуелді болуы және әрі қарайғы аударма алгоритмін құрастыруда ол тәуелділіктің күрделене түсетіндігі.

Көптік аударма жүйесі аталған қатаң тәуелділікте болуына қарсы мүмкіндіктер тудырып, талдау мен жинақтаудың тәуелсіз алгоритмдерін құрастыруға жол ашты. Бинарлық аудармалардың қосындысы ретінде түсіндірілетін көптік аудармадағы тіл жұптарының саны ***n(n-1)*** формуласы арқылы анықталады. Бұл өрнектегі ***n*** – аудару процесіне қатысатын тілдер жиынының саны.

Көптік жүйелерді зерттеуші ғалымдар талдау мен жинақтаудың баламасы (инварианты) ретінде **аралық-тіл** идеясын (интерлингва) ұсынды (*Машинный перевод, 1957*).

Көптік аударма жасаудағы "аралық-тіл" идеясын іске асыруда ең алғашқы елеулі орынға ие болатын ғалымдардың ішінен Н.Д.Андреевтің аты бөліп айтуға тұрарлық (*Андреев, 1957).*

Н.Д.Андреев жетекшілік еткен зертхана (ЭЛМП) қызметкерлерінің осы бағыттағы көптеген ғылыми еңбектерінің жарияланғаны осы саладағы мамандарға мәлім. Осы кезде Кеңес Одағы мен шетелдердің ғылыми орталықтарымен басқа да ғылыми топтар арасында *аралық-тіл* мәселесіне қатысты ой-пікірлер қызу түрде талқыға түсе бастады.

Н.Д.Андреевтің концепциясы бойынша, *аралық-тіл* ұғымының негізінде табиғи тілдердің табиғатында конгруэнттік сипаттың болмайтындығы жайлы түсінік жатыр. Яғни аралық-тілдің жасандылығының заңды екендігі, сол себепті одан конгруэнтті емес элементтердің (артикльдер және т.б.) алынып тасталуы мен көптеген тілдер үшін конгруэнтті элементтердің (етістіктің үш түрлі шағы, көптік жалғау және т.б.) сақталып қалуы қажеттігі туындайды. Сондықтан аталған мәселеге арналған алғашқы зерттеулерге *аралық-тіл* өзінің құрылымы жағынан табиғи тілге жақын тұрып, соның терминдерімен баяндалады.

Әр тілдің бірліктері бойынша грамматикалық мағыналарының, грамматикалық категорияларының және бәріне ортақ болып келетін семантикалық мағыналарының салыстырмалы сипатының қорытындылаушы қасиеттерінің белгілі болуына қарамай, семантикалық теорияның қолданбалы тұрғыда жеткіліксіз зерттелуі тілшілерге дағдылы тілдік көзқарастан ауытқуына кедергі жасады (*Мельчук, 1959; Автомат. пер., 1967*). Н.Д.Андреев жетекшілік еткен машиналық аударманың тәжірибелік зертханасында (ЭЛМП) ЭЕМ-ның көмегімен 18 шет тілінен орыс тіліне аударатын және аралық-тіл идеясын жария ететін шағын дәрежедегі тәжірибе жүргізілді.

Аралық-тіл идеясын өмірге әкелу мақсатындағы сан алуан әрекеттерге қарамай, көптеген тілдерден бір тілге аудару тәжірибесі машиналық аударма тарихында айтарлықтай өз орнын ала алмады деуге болады. Теориялық тұрғыдан қарастырғанда оның себебі аралық-тілдің грамматика мен лексика терминдерінде сипатталуында деп түсіндіріледі. Шындығында, ол тек қана семантикалық тұрғыда, семантикалық барабарлық түсінігінде көрініс табуы қажет болатын.

Сыншылар аралық-тіл арқылы аударма жасаудың төмен дәрежедегі барабарлығын айта келе, оның аударуға тиісті мәтіннің маңызды сипаттамаларының назардан тыс қалып қоятындығын айтады: мәтін авторының коммуникативтік және прагматикалық ұстанымы, талдаудың актуалдығы және т.б.

Машиналық аударма жүйесіндегі аралық-тілдің жүзеге аспай қалуын лингвистикалық зерттеулердің практикалық тұрғыдағы көзқараспен түсіндіретін болсақ, ол өзгеріс енгізу кезеңдерінің әдеттен тыс күрделілігінен, ал техникалық жағынан алғанда ЭЕМ-ның алғашқы буындарының жад көлемі мен әрекет жылдамдықтары көрсеткіштерінің төменгі дәрежелігі деуге болады. Мұндай кедергілер бинарлы аударманың экспериментті жүйесін құру жағдайында да теріс әсерін тигізді.

Қазіргі кезде аталған кедергілер әртүрлі аралық-тілді іске қосу кезінде өз шешімін табуда. Аралық-тіл мәселесіне қызығушылықтың екінші айналымы – жасанды интеллект жүйесін құрастыруға байланысты. Себебі, тек осы идея бойынша ғана қазіргі кездегі кибернетика саласында зерттеліп жатқан мәтіннің мазмұнын түсінетін интеллектуалды жүйелерге жол ашылды деуге болады. Бірақ бүгінгі таңның өзінде-ақ аралық-тіл арқылы машиналық аударма жұмысын жүзеге асыру кез келген аударма жүйесін құрастырушының арманы, жоғарғы мұраты деуге болады. Солай бола тұра, аралық-тіл мәселесін зерттеу семантикалық теорияның, табиғи тілдің фразалары мен мәтіндерінің семантикалық көрінісін анықтау жағдаяттарының даму процесіне зор ықпалын тигізді.

Машиналық аударманың дамуына **аралық-тіл** концепциясының ешбір күмәнсіз үлесі – аударылған және аударылатын тілдердің қасиеттеріне ғана негізделетін анализ бен синтездің бір-біріне тәуелсіз екендігін мойындау; аударылатын сөйлемнің мән-мағынасының үйлесімділігі мен әмбебаптық берілісін зерттеу қажеттігін мойындау болып табылады.

Машиналық аударма жүйелерін құрастыру аясындағы жүргізіліп жатқан жұмыстардың белсенділігіне қарамай, бірінші кезеңде зерттеушілердің ойлағанындай практикалық нәтиже алынбады. Аударманың тұтынушы сұранысын қанағаттандыра алатын өндірістік жүйесін құру жұмысының нәтижесіз болуының үлкен зардабы байқалды. Мысалы, зерттеушілердің бір бөлігі машиналық аударма мәселесін шешуден бас тартты. Осы бағытта жұмыс істейтін бірнеше топтың зерттеу әрекеттерінің тоқырап қалуына себепші болған АҚШ-тың Ұлттық Ғылым академиясының арнайы комиссиясының әуе соғыс күштерінің машиналық аударма жайлы жасаған талдау қорытындысын бұрмалап түсінудің (экстраполяция) де өз рөлі болды. Осының зардабынан көптеген ғылыми ұжымдар ұйымдастыру мен қаржылық қиындықтарға әкелді (*Кулагина, 1979, 14).*

Мұндай жағдайға душар болу кейбір ғалымдардың скепсистік көзқарасы да, машиналық аударма идеясының жүзеге асу мүмкіндігіне сенімсіздік білдіруі де, әсіресе, ғылыми-техникалық мәтіндерді аудару мүмкіндігіне күмән туғызуы да себепші болды. Ал көркем әдебиет мәтіндері, оның ішінде, әсіресе поэзия мәтініне байланысты машиналық аударма жасауға ең басынан-ақ ғалымдар қарсылық пікір білдірген болатын. Оған шығарманың көркемдік, бейнелік жағын ұстаным етудегі кейбір тілдік заңдылыққа қайшы келу жақтары негізгі себеп болды.

Осындай жағдайға байланысты И.Бар-Хиллелдің көзқарасына тоқтала кетуге болады. Ғалым жоғары сапалы толық автоматтандырылған машиналық аударма мүмкін емес және оны утопия деп санаған. И.Бар-Хиллелдің пікірінше, машиналық аударма кезіндегі негізгі кедергі сөздердің көпмағыналығы және тіл табиғатында ондай екімәнділікті ажырата алатын формалды көрсеткіштердің болмауы. Бұл қиындықтардың шешімі тек әлем туралы білімде жатыр деп тұжырым жасайды (*Бар-Хиллел, 1962).*

Жоғарыда айтылған сенімсіздік пікірлер әлем бойынша машиналық аударма мәселесімен айналысатын топтар санын азайтуға әкеліп соқтырды, екінші кезеңнің

аяғында 100 топтан 38 топқа дейін азайғанын айтуға болады (*Кулагина, 1979, 16-18).*

Бірақ машиналық аударманың болашағына сенімді, дәйекті көзқарастағы ғалымдар қажымай, шаршамай-шалдықпай жоспарлаған ғылыми ізденістерін жалғастыра бергені сала мамандарына мәлім. Енді машиналық аударманың екінші кезеңі басталды. Машиналық аударма мамандары зерттеудің теориялық жағын тереңдете түсіп, формалданған лингвистикалық модельдер жасауға, жүйелі және функционалды грамматика жазу мәселесіне кірісті. Бұл кезеңде семантиканы модельдеу жұмысына арнайы назар аударылды. Бірінші кезеңнің нәтижелері сараланып, болашақтағы прагматикаға бағышталған жүйе құрудың мүмкіндігі анықталды.

Бірінші кезеңде зерттеле бастаған формалды грамматиканың модельдері: теоретика-көптік модельдер, бағыныңқылықты грамматикалардың модельдері, тікелей құрастырушылар, туындату грамматикасы мен семантикалық модельдер және т.б. Екінші кезеңде машиналық аударма мәселесінің зерттелуі өз жалғасын тауып, оның интерпретациясына, яғни түсініктеме берілу жағына көбірек көңіл бөлінді, машиналық аударма мәселесі жеке тілдердің грамматикалық құрылымының азды-көпті толық түрдегі сипаттамаларында да қолданыс тапты. Сонымен бірге, олар автоматты талдаулардың іргелі алгоритмдері бойынша да сыннан өтті деуге болады.

Екінші кезеңге қатысты мынадай жайттарды атап кетейік: семантикалық талдау принциптерінің бұрынғыдан да қарқынды түрде зерттелуі және жоғары сапалы аударма жасауға қатысты талдаулар жасау жүйелерінде семантиканың қатысынсыз мәселенің шешілмейтіндігін толық мойындау.

Соңғы айтылған жайтқа байланысты, яғни семантикаға байланысты: біріншіден, кейбір машиналық аударма жүйелерінде талдаудың айрықша семантикалық деңгейін (деңгейлерін) енгізу арқылы (мысалы, ФРАП), ал екіншілерінде – алдын ала белгілі синтаксистік қатынастар типтерінің кең түрдегі синтаксистік сипаттағы тізбесі арқылы мәселені шешу (мысалы, "мән-мағына – мәтін" моделінде), үшіншісінде – аралық-тіл типіндегі мән-мағына инвариантын ұсыну арқылы және осыған әкелетін семантикалық ережелерді қолдану арқылы және т.б. Машиналық аударма жүйесін құруға байланысты жұмыстардың екінші кезеңде жалғасуымен бірге, ол кезең аударманы автоматтандыруды тәжірибелік сынан өткізу және оны өндіріске енгізу әрекеттерімен қоса сипатталады.

Сөздік пен базалық алгоритмдердің көлемдерін ұлғайту арқылы Джорджтаундық жүйе (GAT) және олардың жекелік түрлері даму үстінде болды (мысалы, SYSTRAN жүйесі). 1976 ж. бұл жүйенің ағылшынша-французша версиясы Еуропалық Экономикалық бірлестік қоғамына тапсырылды. Аталған жүйеде негізі синтаксистік байланысты тағайындайтын талдау тәсілі ескерілген болатын (*Марчук, 1985).* Анализ бен синтездің бір-біріне барынша тәуелсіз болуынан оларды басқа салалас тілдік жүйелерге де қолдану мүмкіндігі туындады.

Өткен ғасырдың 70-жылдарынан бері зерттеліп келе жатқан Гренобльдық орысша-французша аударма жүйесінің (GETA) жетілдірілген түрі логикалық пішіндегі терең синтаксистік құрылымдарды алуды көздейтіні мәлім. Бұл жүйе бойынша семантикалық, синтаксистік, морфологиялық және сөздік деректері арқылы бірнеше өңдеулер (трансформаций) жүргізе отырып, "шығыс" тілдің (выходной язык) үстірт тұрпатының (поверхностная форма) көрінісін анықтауға мүмкіндік туады. Көпдеңгейлік талдау өңдеудің 11 топтамасының реттік тәртібі түрінде құрылады. Мысалы, аналитикалық тұлғаны синтетикалыққа келтіру, фразаны сегменттеу, есім топтарын талдау және т.б. Мұндай әрекеттер антецеденттерді

іздестіру мен қалпына келтірілген түйісу нүктелерінің синтаксистік және логикалық функцияларды анықтауға дейін өз жалғасын тапты. Аталған жүйе ARIANE атты дамыған программалық (бағдарламалық) жабдықталуға ие болды. Мысалы, ол жүйе мен тұтынушы арасындағы әрекетті деректер базасы арқылы алдын ала қарастырумен қатар, процесті басқару мен мәтінді өңдеуде де аса дамыған құралдар қолданыс тапты.

TAUM жүйесінде (Монреаль университеті, Канада, жетекшісі – Р.Китредж) ағылшынша-французша аударма үстірттік (поверхностный) синтез деңгейінде жүзеге асады (*Синтаксический компонент ..., 1981*). Аталған жүйе лексикалық және құрылымдық жағынан өңдеу міндетін атқаратын трансфер топтамасына (блогына) ие. Оған лексемаларды аудару, француз тіліндегі фразаны синтездеуді жеңілдету үшін құрастырушылар таралымының (дерево) кейбір бөліктерін өңдеу, етістіктің шақтық түрін және меңгеру моделін (модели управления) аудару мәселелері қамтылды. Программаны (бағдарламаны) жазу тілі ретінде А.Кольмер құрастырған Q (Quebec) тілі алынған (*Синтаксический компонент..., 1981, с. 92)*. Бұл жүйедегі талдағыштың комбинаторлығы мен икемділігі (талдаудың алғашқы болжамын аяғына дейін сақтау, ережелердің қолдану тәсілдерінің ретін белгілеу және т.б.) аударманың үстірттік синтаксиске бағышталуына мүмкіндік тудырады.

TAUM жүйесі 1968 жыл мен 1980 жыл аралығында белсенді түрде дамып, нәтижесінде 1977 жылдан бастап аталған жүйе жыл сайын 5 млн сөз көлеміндегі мәтіндерді аударуды жүзеге асырып, өндірістік жүйе дәрежесіне дейін көтерілді (*Machine Translation6 1987, p. 12)*.

Чехославакиядағы Карлов университетінің алгебралық және есептеу лингвистикасы тобы П.Сгаллдың ғылыми жетекшілігімен жай сөйлемдерді ағылшын тілінен чех тіліне шектеулі сөздік бойынша аударма жасайтын машиналық аударма жүйесін құрастырды (*Синтаксический компонент..., 1981, с. 84)*. Бұл жүйеде талдау процесі негізгі төрт кезеңнен тұрады: морфологияға дейінгі талдау, морфологиялық талдау, синтаксистік талдау, аудару. Синтаксистік талдау синтаксистік ережелердің әрекет ету аясының кеңейтілген типі ретінде құрылған: алдымен шағын мәнмәтіннің бірліктері талданады және әрі қарай оның көлемі біртіндеп ұлғайып, сөйлем шеңберінен сырт шығатын мәнмәтін көлеміне дейін өседі. Сондықтан мұны әдеттегі "төменнен жоғары қарай" талдау жағдайына жатқызуға болады. Осыған байланысты синтаксистік талдаудың өзі де бірнеше кезеңге бөлінеді. Аталған жүйеге жүйесіз (нерегулярный) синтаксисті, мағынаны дұрыс ашуға қажетті сөйлемде жоқ элементтерді толықтыру ережелерін талдау үшін "авариялық" деп аталатын грамматика енгізілген. Синтаксистік қатынастарға семантикалық түсініктеме беру тәсілі кең түрде қолданыс тапқан.

Шамамен алғанда, 1970 жылдан бастап, SUSY жүйесін құрастыру жұмысы басталады (*Синтаксический компонент..., 1981, с. 90*). Аударма жасау процесі деректерді алдын ала дайындаудан, талдаудан, өңдеуден және синтездеуден тұрады. Аударма алгоритмін жасау әр тілдің грамматикасына тәуелсіз сипаттау принципіне негізделеді, сондықтан жүйеге жаңа тілді қосу әртүрлі жаңа бағдарламаларды құрастыру әрекетін сан жағынан шектейді. Неміс, орыс, ағылшын және француз тілдерін талдау, ал сонымен бірге неміс тіліне қатысты синтездеу жұмыстары жүйелі түрде жүргізілді. Туындау таралымы (дерево) түріндегі "базистік құрылымды" – талдау жұмысының нәтижесі деуге болады. Сөздіктер жүйесі (талдау сөздіктері, семантикалық сөздіктер және аударма сөздіктері) әсіресе талдау процесіндегі семантикалық ақпарат үшін аса маңызды болып келетін кез келген ақпаратты пайдалануға мүмкіндік туғызады. Талдау принципі өзара қисынды түрде

құрастырылған: талдаудың бірінші бөлігі "жоғарыдан төменге қарай" өңдеуден басталады, екінші бөлігі "төменнен жоғарыға қарай" өңдеуден басталады, ал үшінші бөлікте алдыңғы екі нәтиже бір-біріне тіркесіп орындалады ("жапсырылады"). Талдау соңында бұрыс синтаксистік құрылымды өңдеу шаралары қарастырылған.

Машиналық аударамен айналысатын ресейлік топтардың сан жағынан аз болуына қарамай, сөз болып отырған екінші кезеңде олардың зерттеу жұмыстарын аса нәтижелі деп сипаттауға болады:

а) лингвистикалық және математикалық жабдықтау тұрғысынан алғанда, бастапқы машиналық аударма жүйелерінің жақсартылған түрлерінің құрастырылуы;

ә) жаңаша типті жүйелердің құрастырылуы (Бүкілодақтық аударма орталығы, А.И.Герцен атындағы ЛГПИ-дегі "Статистика речи" тобы);

б) автоматтанған талдау жүргізетін жүйелердің қарқынды түрде зерттелуі;

в) машиналық аударма жүйелерін әртүрлі мәндік дәрежеге қатысты тәжірибелік сынау деңгейіне дейін жеткізілуі және бұрын-соңды құрастырылған жүйелердің жетістіктері мен кемшіліктері бірдей дәрежеде сыннан өту нәтижелері осыны байқатады.

Аталған кезеңде КСРО ҒА-ның Қолданбалы математика институтында французша-орысша машиналық аударма жүйесінің жаңаша түрі (жаңаша версиясы) құрастырылды. Француз тіліндегі математикалық мәтінді орысшаға аударатын бұл жүйе 1967–1976 жылдар аралығында О.С.Кулагинаның ғылыми жетекшілігі бойынша зерттеу жұмыстарын жалғастырды (*Кулагина, 1979).*

Жүйеде аударма әрекетіне қатысты өңдеу жұмыстары былайша іске асты: жүйеге "кіретін" ақпарат французша фразаның үстіңгі синтаксистік құрылымы (фраза бойынша аудару), ал жүйеден "шығатын" ақпарат – мән-мағыналық сәйкестікте тұратын орысша фразаның үстіңгі (поверхностный) синтаксистік құрылымы. Мұндай тәсіл екі түрлі тілдердің сөздері мен фразалары мағыналарының сәйкестігі мен олардың құрылымдық сәйкестік сипатына негізделеді. Бұлайша талдаудың негізгі мақсаты бағыныңқылықтар таралымы (дерево) түріндегі фразалардың синтаксистік құрылымдарының жиынын құру. Осылайша талдау әдісін жүргізе отырып, синтаксистік байланыстың 52 типі жеке түрде бөлініп алынады.

Көпвариантты талдауда сүзбе-ережелер жиыны қолданылады. Мұндай әдіс арқылы синтаксистік байланыстың күші мен арақашықтығы және байланыстың ортақ немесе ортақ еместігі (совместность/несовместность связей), құрылымдардың жобалауға болатындығы немесе болмайтындығы (проективность/ непроективность структур) ескеріледі.

Сүзбелерді қолданудың ретіне қарай, олардың иерархиялық айырым-белгілері есепке алынған. Мысалы, алдымен қарапайым сүзбе, сосын барып жалпы, күрделі, ал ең соңында жекеленген жағдайға сай келетін сүзбе түрлері қолданылған. Көпвариантты талдау түрлеріне сүйене отырып, болжамдар жиыны ішінен аталған сүзбелер арқылы нақты бір болжамды құрастыруға мүмкіндік туындайды (*Кулагина, 1979; Синтаксический компонент..., 1981*).

Орыс тілін синтаксистік талдау арқылы зерттеу жолында машиналық аударманың алғашқы кезеңінде орын алған конфигурациялық талдау идеясы қолданыс тапты. Осының нәтижесінде сөйлем ішіндегі сөздердің әр алуан синтаксистік қатынас типтерін мағыналық тұрғыда сипаттайтын **тікелей доминация қатынасы** (отношение непосредственной доминации) деп аталатын 31 қатынасы жеке бөлініп алынды. Мұндай синтаксистік талдаудың негізгі мақсаты аударылатын сөйлемнің сөздері арасындағы синтаксистік байланысты анықтау болып табылады. Ол үшін мұндай байланыстарға қажетті тікелей доминация қатынастарын тіркеп

жазу және оның мазмұнына қатысты ерекше сипаттамалар сәйкестендірілуі қажет болады. Автоматты талдау жүйесінің мәтінді аудару, реферат алу, ақпаратты іздестіру сияқты қолданбалы мәнімен бірге, теориялық жағы да аса құнды деуге болады. Себебі, талдау алгоритмі "тілдегі талдаушы бөліктің" әрекет ететін және эксперименталды түрде бақыланатын моделі болып табылады (*Автоматический перевод, 1967. 226; Мельчук, 1974).*

Кезінде Ленинград университетінің математикалық лингвистика зертханасында (НИИММ) ағылшынша-орысша аударма жүйесі тәжірибелік тексеру дәрежесіне дейін жетті (*Лейкина және т.б., 1966).* Бұл жүйеде талдау мен жинақтауды бір-біріне тәуелсіз қарастыру мүмкіндігі есепке алынды.

Ал өңдеу мен түсініктеме беру әдістерінің міндеттері – бірізділікке келтіру. Талдау әдісі фразалық және көптік сипатта жүргізілді. Бағыныңқылық грамматикасы моделінің шеңберінде тұжырымдалған лингвистикалық мәліметтер, оларды өңдеу шараларынан (алгоритмдерден) жеке қарастырылды. Ал алгоритм жұмысының нәтижелігі синтаксистік талдау әдісін дұрыс жүргізу, жобалау, тарамдалу (древесность), байланыстың міндеттілігі, ортақтығы және т.б. шарттарын қамтамасыз ету арқылы жүзеге асты.

Көп ұзамай-ақ, жоғарыда аталған зертханада байланыстың тұйықталу сипатына айрықша мән беру мен реттілікті есепке алатын орыс тілінің синтаксистік жүйесін құру және оны тәжірибелік сынақтан өткізу жұмысы жүргізіледі (*Кремнева және т.б., 1976).*

Ю.Д.Апресянның жетекшілігімен ИНФ-ЭЛ атты лингвистикалық жағынан аса күшті жабдықталған французша-орысша аударма жасау жүйесі құрылды (*Синтаксисческий компонент..., 1981).*

Аударма үстіңгі (поверхностный) синтаксистік құрылым деңгейінде жүргізіліп, мына принциптерге негізделді:

1) аударылуға тиісті тілдердің грамматикаларының сипатталуы "мән-мағына–мәтін" ("смысл-текст") атты бірыңғай модель негізінде құрылуы;

2) талдау мен жинақтау ережелерінің бір-біріне ауысу мүмкіндігіне және олардағы ақпарат көлемінің жеткілікті болуы;

3) грамматика мен сөздікке қатысты сипаттаулардың алгоритмдер мен программалардан бөлектік сипаты (аударманға қатысты ережелерден басқа);

4) аударуға қатысты тілдердің сипаттауларының тәуелсіздігі;

5) әртүрлі пәндік салаларға бейімделуінің әмбебаптылығы;

6) лингвистикалық білімді бейнелеу үшін формалды аппаратты (модель және тіл) құру.

Бірінші кезекті ИНФ-ЭЛ жүйесі "ИРИС-50" ЭЕМ-да жүзеге асты.

Мәскеудегі М.Торез атындағы шет тілдер институтында АРМАС атты машиналық аударма жүйесі үшін көпаспектілі ағылшынша-орысша сөздік құрастырылды. Оның мақалалары сөздік бірлігінің әр алуан лингвистикалық сипаттамаларын қамтиды: морфологиялық, синтаксистік, семантикалық, лексикалық қызметі жайлы мәліметтер, меңгеру моделіне қатысты аударма баламалары жайлы және т.б. (*Шаляпина, 1974).*

Бұл сөздік АРАП атты ағылшынша-орысша автоматты аударма жүйесінің лингвистикалық жабдықталуының маңызды бөлігі ретінде түзілді (жетекшісі – З.М.Шаляпина). АРАП жүйесі үшін қабылданған ұстаным ЯРАП атты жапонша-орысша автоматты аударма жүйесі үшін де қолданылды.

Аталған жүйелер үшін талдау жұмысы семантика-синтаксистік деңгейде жүргізілді. Тілдік бірліктердің семантикалық тіркесін зерттеу мақсатында арнаулы

тіл құрастырылып, барлық негізгі лингвистикалық ақпараттар біртиптік сөздік мақалалары түрінде көрініс тапты. Оларда сөздік бірліктері грамматикалық, лексикалық, тіркесімдік және басқа да әр фразаны аударуға жеткілікті белгілер түрінде көрініс тапты.

Бүкілодақтық аударма орталығы жұмысының екінші кезеңінде бірнеше аударма жүйелері құрастырылды. Олардың ішінен тек екеуін ғана атап өтейік: Н.Н.Леонтьева жетекшілік еткен ФРАП атты французша-орысша автоматты аударма жүйесі (*Леонтьева, 1987*) және Ю.Н.Марчук жетекшілік еткен АМПАР атты ағылшынша-орысша автоматты жүйе (*Марчук, 1983).*

ФРАП жүйесі аудармашының әрекетін модельдеуге арналған аса қызықты талпыныс деуге болады. Аударылатын мәтін тілінің аудармашының ана тіліне қатысты еместігі ескерілді, сондықтан мәтін ішінде аудармашыға белгісіз тілдік элементтердің кездесетіндігі де есепке алынған. Ғылыми тұрғыдағы мұндай зерттеудің мақсаты – таза грамматикалық тәсілдің "күшін" анықтау. Осыған байланысты аударма процесі қажетті фразаны талдау деңгейінің белгілі сатысына сәйкес келетін көптеген топтамаларға (блоктарға) бөлінеді. Ең алдымен формалды-грамматикалық құралдардың барлық мүмкіндіктері пайдаланылады, ал содан кейін ғана талдаудың семантикалық деңгейі іске қосылады. Семантикалық деңгейдің грамматикалық деңгей тәріздес, сан жағынан бірнешеу болуы алдын ала ескерілген. ФРАП жүйесіндегі аударманың концепциясы жүйенің синтаксистік компонентінің "күшін" бағалайды. Бұл жүйенің ерекшелігі – аса дамыған семантикалық аппаратпен әрекет етуі. Бірінші синтаксистік топтаманың жұмысы аяқталғаннан кейін ғана келесі семантикалық аппаратқа бетбұрыс басталады. Жүйенің әр деңгейіндегі ішкі кішігірім жүйелерді тәжірибелік сынақтан өткізу де алдын ала қарастырылған.

Грамматикалық және дамыған лексикалық талдауларға бағышталған АМПАР жүйесі де көпдеңгейлі. Онда сөз айналымдары, бірмағыналы және көпмағыналы сөздер жеке топтамалар арқылы өңделіп, тек сөйлем құрылымдары ғана өзгеріске ұшырайды. Аталған жүйе іс жүзінде қолданыс табуда.

Белгілі ғалым, профессор Р.Г.Пиотровский жетекшілік ететін Бүкілодақтық "Статистика речи" тобы 1970 жылдан бастап машиналық аударма жүйесі бойынша зерттеулер жүргізді. Бұл жүйе қазіргі уақытта да актуалды болып саналатын итерактивті-стратификациялық тәсілге негізделеді. Мұндай әдіс бойынша аударуға тиісті мәтінді қайшылықсыз сипаттау үшін ол кезеңдік сипатта қарастырылуы қажет.

Ленинградтағы А.И.Герцен атындағы Педагогикалық институттың инженерлік лингвистика зертханасында 1975 жылдан бастап зерттеле бастаған СИЛОД атты аударма жүйесі көптілді жүйе ретінде өз міндетін атқара бастады. Оның сөздіктермен және морфологиялық тұрғыда жабдықталуы аса күшті және құрамында трансфер қолданыс табатын грамматикасы бар жүйе болып саналады.

Осы кезеңдерде ағылшынша-орысша және французша-орысша машиналық аударманың түрлері құрастырылып, тәжірибелік сыннан өтеді. Осы әрекеттердің нәтижесінде жекеленген синтаксистік конструкциялардың құрылымы мен мағынасын тануға мүмкіндік тудыратын талдаудың фреймдік және тезаурустық негіздері қаланады (*Пиотровский, 1979; Чижаковский, Беляева, 1983).* Аталған жүйеде машиналық аударманың алғашқы кезеңінде ұсынылған "фулькрумдар" әдісінің дамыған түрі ретінде саналатын машиналық аударма алгоритмдерін деңгейлік тұрғыда құрастыру ескерілген.

Машиналық аударма жүйелерін құрастырудың екінші кезеңін қысқаша қарастыра отырып, олардың негізі сипаттамалары мен жалпы ерекшеліктерін былайша атап кетуге болады (*Кулагина, 1979; Hutchins, 1986)*:

1) көпдеңгейлі талдау мен синтаксистік деңгейдің толық дәрежеде көрініс табуы;

2) талдау нәтижелерінің көпвариантты болуы. Аталған жайт көптеген мүмкіндіктердің ішінен ең дұрысы деп саналатын біреуін ғана таңдап алу ықтималдығын күшейтеді.

3) лингвистикалық мәліметтерді сақтайтын қатал алгоритмнен барынша "жұмсақ" жүйеге ауысу мүмкіндігі. Соңғы жүйеде барлық лингвистикалық мәліметтер олардың алгоритмінен бөлектелген (*Лейкина және т.б. 1986; Кулагина, 1979).*

Жоғарыда аталған жүйелердің сипаттамалары мен ерекшеліктері машиналық аударма жұмысын ұйымдастыруға оң әсерін тигізді. Басқаша айтқанда, олар тілшілер мен математика мамандарына жүйенің ерекшелік аспектілерін қарқынды түрде зерттеуге бағыт-бағдар беріп, ынталандыруға себепші болды. Сонымен бірге аталған жайт тілдің формалды моделін құру жолын зерттеу мен оны дұрыс қолдану мүмкіндіктерін анықтауды қажет етті. Ал табиғи тілді формалды сипаттау әдісі грамматикаларды бірізділікке келтіруге мүмкіндік тудырды. Сондықтан машиналық аударманың даму кезеңінің екінші сатысында әртүрлі тілдерді сипаттау үшін формалды модельдер қолданумен бірге тілдік деректерді қалаған дәлдікпен (адекватно) сипаттауға мүмкіндігі мол жаңа модельдер үлгісін құру қажет болды.

Сонымен, машиналық аударманың даму кезеңінің екінші сатысы бірінші буынға қатысты практикалық жүйелердің дамуымен, тәжірибе жинақтаумен, оны іске қосумен, сөздік базасының кең түрде өріс алуымен және тәжірибелік тұрғыда болса да барынша жабдықталған теориялық жүйелерді құрумен сипатталады.

Қол жеткен практикалық және теориялық нәтижелер машиналық аударманың даму сатысының келесі үшінші кезеңіне көшуге даяр екендігінің айғағы болып саналды. Бірақ келесі сатыға өту жолы бірнеше экстралингвистикалық деректерді алдын ала анықтап алуды қажет еткені белгілі.

Біріншіден, 70-ші жылдардың аяғында ақпарат ағымы бұрын-соңды болмаған қарқында дамиды. Мысалы, американдық және канадалық корпорациялар бір жылдың ішінде 2,5 млрд сөзге дейінгі мәтіндерді аударып отыруға мүмкіндігі болған. Егер АҚШ-та 1966 ж. аударма жұмысының біржылдық шығыны 22 млн доллар болса, 1981 ж. ол шығын 200 млн долларға дейін өседі. Кеңес Одағында штаттағы аудармашылардың айлығын санамағанда, тек тапсырыс бойынша орындалатын аударма жұмысының өзіне орасан үлкен қаржы жұмсалған екен. Бірақ ақпарат ағымының (ғылыми-техникалық мәтіндер, патенттер, құжаттар мен арнаулы қағаздар) соншалық үлкен дәрежеде болуына байланысты оларды тек дәстүрлі жолмен, яғни аудармашылар күшімен аударып шығу мүмкін емес еді. Сол себепті, қазіргі кезде автоматтандыру арқылы техникалық стильдегі мәтіндерді аударуды жеңілдету мен жетілдіру маңызды мәселеге айналып отыр.

Екіншіден, спеллер типті өндірістік лингвистикалық процессорларды құрастыру мен оларды баспа мен лексикографиялық жұмыстарда іс жүзінде қолдану, сол сияқты сөздерді өңдеу жүйесінде (word processors) күшті терминологиялық банктерді пайдалану – "лингвистикалық" автоматтанған жүйелерді қолдану арқылы нақты практикалық нәтижелерге ие болу мүмкіндігінің кепілі.

Үшіншіден, машиналық аударманың жаңа кезеңіне өтуге қатысты ерекше жасалған алғышарт ретінде есептеу техникасының дамуына байланысты барынша жағдай жасалуы – дисплейлердің пайда болуы, дербес компьютерлер мен сканерлердің өмірге келуі, дербес ЭЕМ-ның күш-қуатының өсуі және т.б. Міне, осының бәрі теоретика-лингвистикалық зерттеулерге көзқарасты басқаша өзгертіп, оны практикалық деңгейге дейін көтерді.

Машиналық аударма жүйелерін құрудың бірінші және екінші кезеңдерін сипаттауда, біз нақты лингвистикалық ұжымдардың жұмыстарына айрықша назар аудардық. Ал енді машиналық аударманың қазіргі кезеңін сипаттау үшін оның стратегиясын таңдау мен практикалық жүйелерін тиімді пайдалануға байланысты негізгі теоретикалық және техникалық мәселелерді анықтауға қатысты кейбір мәселелерге тоқталайық. Әрине, машиналық аударманың мұндай негізгі мәселелерін талқылау машиналық аударма жүйесінің болашақтағы даму жағдаятын да сөз етуді қажет етеді.

Бірінші мәселе мынадай мүмкін болатын үш түрлі теориялық стратегияларды талқыға салуды қажет етеді:

- тікелей бинарлық машиналық аударма;
- трансфер арқылы аудару;
- аралық-тіл арқылы аудару, сонымен бірге олардың грамматикалық мүмкіндіктерінің ерекшеліктерін іске асыру.

Тікелей бинарлық аударма машиналық аударманың "ескі" жүйелерінен орын алса да, сол қалыпта машиналық аударманың көптеген практикалық жүйелерінің негізгі стратегиясына айналып отыр. Мұндай тәсіл нақты тілдік жұпқа, яғни қос тілдерге бағытталады (GAT, SYSTRAN жүйелері және олардың дереваттары Бүкілодақтық аударма орталығында зерттелген АРМАС жүйесі). Бинарлық аударма әрекеті кезінде сегменттердің шекарасын анықтау мен омонимияны мәнмәтінге қатысты ажырату негізінде сөйлемдерге қарапайым өзгертулер (трансформация) жүргізіледі.

Бұрынырақ зерттелген машиналық аударма жүйесін құрастыру тәжірибесі мен оны пайдалану тиімділігі, олардағы аса күшті сөздіктер қоры мен арнайы түзету (редакциялау) тәсілдерінің негізінде деуге болады. Оған Джорджтаун университетінің GAT жүйесі мен оның түрлі вариантық жүйелері мысал бола алады. Бірақ жеке лексика-морфологиялық және семантика-синтаксистік ішкі құрылымдар (подструктуры) негізінде іске асатын тікелей бинарлық сәйкестендіру жүйенің лингвистикалық және программалық жабдықталуын тек нақты тілдер жұбының ерекшеліктеріне ғана бағыштап қоймай, сонымен бірге ішкі тілдердің (подъязык) айырым-белгілеріне де қатаң түрде көңіл бөлуді қажет етеді. Егер тіл жұптарының бір сыңары басқа бір тілге ауыстырылса, іле-шала лингвистикалық, ал кейбір жағдайларда программалық жабдықтауларды да қайта құру қажет болады.

Трансформациялық машиналық аударма (трансфер арқылы аударма) әдісі де машиналық аударма дамуының екінші кезеңінде қолданылып, сынақтан өтті. Трансферді пайдаланатын бұл стратегияны қолдану – түрлі тілдердің мәтіндеріне тікелей қатысы бар жазбаларды жүзеге асырудың мүмкін еместігі жөніндегі теориялық болжамға негізделеді. Мұндай байланыс түрі әртүрлі деңгейдегі жалғас жұмсалымдар жолымен анықталуы мүмкін: лексикалық, морфологиялық, синтаксистік құрылымдар деңгейінде (конфигурациялық және реляциондық синтаксис терминдерімен) және синтаксис деңгейінде сипаттау. Түрлі деңгейдегі талдаулар кезінде "кіретін" сөйлемнің құрылысын сипаттайтын трансферді пайдалану бағыныңқы таралымнан базалық грамматикалық бірліктер ретіндегі басты астырт құрылымға (к глубинной ролевой структуре) біртіндеп өтуді көздейді және одан әрі "шығатын" бағыныңқы таралымға өтеді. Мұндай кезде тек белгілі бағыныңқы таралымға қатысты қайта құру әрекеті сөйлемнің толық түсінігінің моделін құрастырмай-ақ машиналық аударма жасауға мүмкіндік тудырады. Бұл мәселенің шешімін табу машиналық аударма кезінде міндетті емес, өйткені астырт (глубинной) семантикалық-синтаксистік құрылымдарға түсініктеме беру

тұтынушының үлесіне қалдырылады, яғни ЭЕМ-нан шыққан аудармамен әрі қарайғы жұмыс істейтін маман аудармашының үлесіне тиеді. Трансферді қолданудың мұндай тәсілі көптеген шетел ұжымдарымен бірге машиналық аударма жүйесін қолданатын бірқатар ресейлік ұжымдарға да тән болып келді.

Аталған жүйелерде аударманың негізгі үш сатысын бөліп айтуға болады: *талдау* (анализ), *трансфер* және *жинақтау* (синтез). Әрбір саты жүйелі субпроцестер түрінде орындалады. Мәселен, жоғарыда аталған SUSY атты жүйеде талдау әрекеті 8 субпроцесті қамтиды. Нәтижесінде талдауға қатысты "кіретін" сөйлемнің таралымы құрастырылады және "шығатын" сөйлемнің таралымы жайлы мәлімет 8 субпроцессордан тұратын трансфер түрінде көрініс табады. Талдау кезеңінде бұл таралым үш субпроцестің өзара әрекеттесуінің нәтижесінде аралық-тілдегі сөйлем түріне қайта келтіріледі. Осыған ұқсас тәсілдер трансфермен әрекет ететін басқа да жүйелер түрлерінде қолданылады: CADA, GETA, TAUM-METEO, жапон жүйесі, EUROTRA жүйесі, Ресейдегі СИЛОД (MULTIS) жүйесі, А.И.Герцен атындағы РГПУ-да зерттелген ФРАП, АРАП, ЭТАП, ЭТАП-2, STYLUS және т.б.

Еуропалық экономикалық бірлестіктің айрықша комиссиясы жанында 1982 жылы машиналық аударманың дамуының үшінші кезеңінде ұсынылған EUROTRA жүйесінің негізі қаланды. Аталған жүйенің тілдері: ағылшынша, французша, немісше, итальянша, испанша, португалша, голландша, датша, жаңагрекше. Бұл жоба БЭЖ (ЕЭС) елдерінің тоғыз ресми тілі үшін машиналық аударма жүйесін құруды және 72 тілдің жұбы бойынша аударма жасау қарастырылды.

Аталған тілдердің қайсысын болса да аударуға қажетті талдауда интерфейстік көрініс деп аталатын синтаксистік қатынастардың семантикалық түсінігі беріледі. Мәтін семантикалық бағыныңқылықтардың түзілістері арқылы көрініс табады. Мысалы, олар астырт септігі типінде (типа глубинных падежей) болуы мүмкін. Бұл жерде арнайы атап кетуге тұрарлықтай жайт – бір тілдің интерфейстік құрылымының басқа тілдегі осындай құрылыммен сәйкес келмеуі және осыған қарамай трансферді қолдануға қажетті мүмкіндіктің туатындығы. Сонымен, ана тілінің аясында аударма жасайтын әрбір маман талдау мен жинақтау тәсілдерінен басқа трансфердің 8 түрлі процедурасын да құра білуі қажет.

Жобаның мақсаты – өндірістік машиналық аударма жүйесінің бастапқы үлгісін жасау. EUROTRA жүйесінде іске асатын трансфердің ерекшелігі, оның қадамдап әрекет етуі. Әрбір қадамды жасаудан бұрын алдыңғы қадамдағы әрекеттердің бұрыс нәтижелері "тазартылады".

Трансферді пайдаланатын жүйелерге ЭТАП-2 атты жүйе де жатады (*Апресян и др. 1989).* Бұл жүйе ағылшынша-орысша машиналық аударма жасай алады. Бұл идеологиялық тұрғыдан қарастырғанда ИНФ-ЭЛ жүйесінің жетілген түрі және айрықша сөздіктермен жабдықталған – нақты бір пәндік салаға арналған, ұластыруға негізделген (комбинаторлық) автоматты сөздік. Зерттеушілер мұндай "тәуелсіздік" мүмкіндігін тілдік модельдің жан-жақты зерттелуімен байланыстырады.

Трансфердің барлық ережелері жүйе ішінде **жалпылық**, **жекелік** (трафареттік) және **сөздіктік** деп үшке бөлініп қарастырылады. Жүйеде 45 жай (элементарлық) предикаттар мен 5 синтаксистік қатынастар орын алған. Сүзбеден өткізу тетігі (механизмі) мен жоғары баға беру ережелері ұсынылған синтаксистік болжамдар ішінен тек біреуіне ғана артықшылық беруге мүмкіндік туғызады.

Трансферді қолдану талдаудың әртүрлі деңгейінде "кіретін" сөйлемнің құрылымын сипаттайтын ішкі тармақтардан (поддеревья) базалық грамматика берілісін қолданатын астырт басты құрылымға (к глубинной ролевой структуре) жүйелі түрде көшу жолы мен одан әрі қарай "шығатын" ішкі таралымдарға көшу

әрекеті қарастырылған. Осындай жағдайда белгілі ішкі тармақтар үшін қайта құру әрекеттері сөйлем түсінігін модельдемей-ақ, машиналық аударма жасауға мүмкіндік тудырады. Ұғынуды модельдеу міндеті машиналық аударма мәселесінен тыс жатады, өйткені машиналық аударма мәселесіндегі астырт (глубинный) түсініктеме алу адам үлесіне, яғни нақты пәндік сала маманына жүктелетіні жоғарыда да аталды.

Трансфер арқылы машиналық аударма жүйелерін іске қосатын мұндай тәсіл машиналық аударма жүйесі дамуының екінші кезеңінде пайда болған СИЛОД жүйесінде де қолданыс тапты. Қазіргі кезде SILOD-MULTIS көптілді жүйе дербес компьютерлерде аударма жұмысын атқарады. Машиналық аударманың барлық процесі лексика-морфологиялық және семантика-синтаксистік субпроцестердің композициясы ретінде SILOD-MULTIS жүйесінде модельденеді. Лексика-морфологиялық талдау кезінде мәтіндер мен сөздік бірліктерін сәйкестендіру жұмысы жүргізіледі және сонымен бірге автоматты сөздік топтамасынан сөздікке қатысты ақпараттар шығарылып алынады. Мұндай талдаудың нәтижесі ретінде мәтіннің кесте түріндегі көрінісін және канондық пішіндегі аударма баламалары (переводные эквиваленты) жайлы ақпаратты айтуға болады. Бұл кестеде әрбір сөзтұлғаға оның лексика-семантикалық сипаттамасы сәйкестікке қойылады. Сәйкестіктің дәлдігі конверсиялық және/немесе омонимдік септелуге (подежная омонимия) дейін, функционалдық, басты (ролевой) және семантикалық сипатына дейін анықталады. Мәтінді талдаудың барлық деңгейлеріндегі берілісі сияқты, аталған көріністер әртүрлі жағдайда қолданылуы мүмкін (машиналық аудармада, аннотациялауда, рефераттауда және т.б.). Егер машиналық аударма жайлы ғана айтатын болсақ, онда талдаудың бұл сатысында алынатын нәтиже – лексика-морфологиялық және фразеологиялық тұрғыда жолма-жол (подстрочник) аударылған мәтін.

Семантика-синтаксистік өңдеудің нәтижесінде құрылымды тіларалық деңгейде қайта құру жүзеге асады. Ол үшін лексика-морфологиялық талдау сатысындағы барлық ақпарат іске қосылады, одан кейін барып грамматикалық және семантикалық топтамалар пайдаланылады. Мұндай трансферде иерархиялық бөліктер (компоненттер) мәтін ішінен алынатын тік қисындасқан жүйешіктердің жиынтығы ретінде модельденеді.

Семантика-синтаксистік талдау сөйлем құрылымының әртүрлі деңгейлерінде жүйелі түрде жүзеге асады: топтар деңгейі, функционалды сегменттердің деңгейі және, ең соңында, сөйлем деңгейінде (*Беляева, 1986; Апресян және т.б., 1989, Соколова, 1990).* Топтар деңгейінде топтардың шекаралары мен лексикалық бірліктің өзара байланысы белгіленеді. Келесі деңгейде бөлініп алынған тізбектер сөйлемнің тұрлаулы мүшелеріне сәйкес келетін функционалды сегменттерге біріктіріледі. Бөлініп алынған сегменттерде мәтінді кестелік көрініске модификациялау деңгейінде қайта құру әрекеті жүзеге асады. Функционалды сегменттер деңгейінде алынған нәтиже – шектеулі грамматика бойынша жасалған машиналық аударма болып саналады. Толық түрдегі семантика-синтаксистік талдау сөйлем деңгейінде ғана іске асады.

Талдау мен трансфер процедураларын жүзеге асырғанда синтаксистік сипаттау маңызды рөл атқарады, өйткені тек осының негізінде ғана сөйлемнің мүмкін болатын фреймі танылады. Мұндай таным әрекетінде актанттар жиыны, меңгеру моделінің мүмкіндігі жайлы және т.б. ақпараттарды бойында сақтайтын етістіктің лексика-синтаксистік сипатталуы негізгі рөл атқарады.

Осы әдістемеге сәйкес SILOD-MULTIS машиналық аударма жүйесі автоматты сөздіктер мен грамматикалардың жинағы ретінде жүзеге асады. Жүйенің негізгі

пішін үйлесімділігінде үндіеуропа және басқа құрылымдық тілдермен әрекет ету мүмкіндіктері алдын ала қарастырылған. Мұндай мүмкіндік барлық "кіретін" тілдердің сөзтізбелерінің әмбебаптық құрылымына және "шығатын" орыс тілінің бірыңғай құрылымына негізделеді.

Нақты әрекеттегі машиналық аударма жүйелерінің трансферді пайдалануы дәл осы стратегияны таңдаудың мақсатқа сай екендігін дәлелдей түсті. Сонымен бірге жасанды интеллект жүйесінің даму деңгейі мен есептеу техникасының жаңа мүмкіндіктері машиналық аударманың үшінші стратегиясына қызығушылықты жаңаша түрде жандандыра түсті.

Аралық-тілді қолдануға қатысты аударма стратегиясы машиналық аударма прагматикасына білім базасын құрудағы машиналық аударма мен оның идеологиясының әсерін бейнелейді. Машиналық аударманың даму сатысының бірінші кезеңінде қолданылатын аралық-тіл тәжірибе жүзінде ғана жүзеге асқанымен, ол өз уақытынан айтарлықтай озды – оның пайда болуы машиналық аударманың жаңа кезеңі деуге болады. Мысалы, аралық-тілді қолдана отырып, Карнеги университетінде CMU жүйесі, Утрехте зерттелген DLT жүйесі, Японияда – JETR, АҚШ-та – SAM және TRANSLATOR жүйелері машиналық аударма жасауда қолданыс тапты.

Аралық-тілді пайдаланатын машиналық аударма жүйесі мен жасанды интеллект жүйесін бір-біріне қисындастыру мәселесі – сөйлем мағынасын немесе тұтас мәтіннің мазмұнын тануға және оның мазмұнын автоматты түрде сөйлеу тілінде туындату үшін күрделі білім базасын қолданудың қажеттігімен түсіндіріледі.

Жасанды интеллект жүйесінің аралық деңгейі болып саналатын ұғынудың (понимание) концептуалды деңгейін модельдеу бірнеше базалық құрастырушыларға негізделеді:

- ұғым түйіндері (узлы понятия) мен байланыс типтері жайлы сөз болатын мақала түріндегі сөздіктерді құрастыруды қажет ететін білімді тануға көмектесетін тілдер;
- "белгі – мағына" принципі бойынша қайта құруды жүзеге асыратын жүйелі грамматикалар.

Қазіргі кезеңдегі аралық-тілдің негізі ретінде білімді танытатын тілдер саналады. Осыған қатысты талдаудың мақсаты "кіретін" сөйлем бойынша алынатын және білім базасының ақпаратымен толықтырылатын сөйлем мағынасын таныту. Аударманы жүзеге асыру әрекетінде "кіретін" тілдің құрылымын "шығатын" тілдің құрылымына сәйкестендіріп, қайта құру қажет болады. Әрине мұндай аударма нәтижесі шын мағынасындағы аудармадан гөрі, берілген мәтіннің мазмұнын не сөйлемнің мағынасын қайталап айтып беру жағдайына көбірек ұқсайды.

Осындай жағдаятта білімді танытатын тілді аралық-тілдің өзі деуге болады. Мұндағы аралық-тіл "кіретін" фразаның (сөйлемнің) семантикалық желісі мен "шығатын" фразаның ішкі берілісінен қайта туындауға қатысты семантикалық таңбалауды ("разметка") жүзеге асырады. Осылайша қайта туындату (генерация) белгілі үлгі бойынша (шаблон, фрейм және т.б.) орындалады. Бұл әдістің жалпытеориялық тартымдылығына қарамай, ол тек тәжірибелік деңгейде ғана жүзеге асты. Себебі оның автоматтандырылған лексикондарының негізінде аса күшті семантикалық желі құруға қажетті лингвистикалық негіздемесі мен программалық құралдары әлі де болса толық қамтылмаған еді.

Машиналық аударма жүйелерінің кең түрде қолданыс табуы оның эксперименттік дәрежесінен автоматтанған дәрежесіне, редакциялау ерекшелігіне, мақсатқа сай келуіне қарай аудармашының жұмыс орнына айналуына жағдай жасайды (*Machine Translation, 1987; Апресян және т.б., 1989).*

Машиналық аударма жүйелерінің қазіргі сатыдағы даму ерекшеліктерін қарастыра келіп, оны әр кезеңге бөліп топтап, баға беру дұрыс бола бермейді. Себебі машиналық аудармаға қатысты ондай топтаулар есептеу машиналарының (компьютерлердің) кезеңді өзгерістеріне және мүмкіндіктерінің артуына тікелей байланысты.

Қазіргі кездегі машиналық аударма жүйелерінің зерттелуі мен қолданылуы мына жайтты байқатады. Машиналық аударма жүйелерін талдау және олардың функционалдық критерийлері бойынша салыстыру мынаған саяды:

1. **Жүйенің прагматикасы.** Машиналық аударма жүйесін өндірістік немесе эксперименттік дәрежеде құруға байланысты оның негізгі сипаттамасы мен талдау әдістемесін таңдау түрлері ерекшеленеді. Бұл өлшем сөздік базасының көлеміне қойылатын талаптарды (өндірістік машиналық аударма жүйесінде бір тілдер жұбына 20000 сөздік мақаладан кем болмау) анықтайды. Мысалы, Bravice – Weidner LOGOS, GAT, S'YSTRAN және т.б. жүйелерде сөздік мақала көлемі көрсетілген мөлшерден кем емес. Бұл шартқа бағынбайтын жалғыз жүйе – TAUM-METEO. Оның сөздік мақала көлемі 1500-ге тең, өйткені бұл жүйедегі "кіретін" мәтіндердің тақырыптары қатаң түрде шектелген. Тәжірибелік (эксперименттік) жүйелерде автоматты сөздіктер "төменгі" жағынан шектелмеген, яғни көлемі 200 сөздік мақаладан тұратын сөздіктер бойынша да ол жүйе аударма жұмысын атқара алады.

Өндірістік машиналық аударма жүйесін құрудың мақсаты анық – үзіліссіз процесс түрінен ұйымдастырылған құжат ағымдарынан машиналық аударма алу.

Машиналық аударманың өндірістік жүйелері мынадай талаптарға сай келуі керек:

- **Жүйе жұмысының тұрақтылығы.** Машиналық аударма жүйесінен алынатын нәтиже сапасының ойдағыдай болмауына қарамай, тұтынушы оны өз қажеттігіне пайдалана алатындай мүмкіндігі болуы керек. Машиналық аударма жүйесінен "шығатын" ақпараттың сапасының төмен болуы, көбінде, жүйенің сөздіктермен толық жабдықталмауына байланысты болып келеді. Сондықтан жүйеден "шыққан" кейбір сапасы төмен ("дөрекі") аудармалар әрі қарай маман аудармашының қолынан қайта түзетіліп, сапасы жақсартылады. Осыған байланысты SYSTRAN жүйесін пайдалану тәсілімен танысу қажет.

- **Жүйенің тираждануы (таратылуы).** Аударма жүйесінің қолдану аясы кең болуы үшін оның программалық және лингвистикалық құралдары барынша қарапайым түрде болуы қажет. Сонымен бірге жүйеде басқа тақырыптық аяға және тілдерге көшу әрекетін жеңілдетуге қажетті әмбебаптық модельдер орын алуы керек.

- **Жүйенің икемделінуі (адаптациялануы).** Машиналық аударма жүйесінің нақты тұтынушының мұқтаждығына және өңделетін материалдың ерекшелігіне қарай икемделу мүмкіндігі болуы қажет.

- **Уақытша параметрлердің оптималдығы (ауқымдылығы).** Машиналық аударма жүйесінің аударма жасау жылдамдығы уақыт бірлігі ішінде қабылданатын ақпарат көлемімен немесе жүйені тұтынушы ретіндегі редакторлардың, аудармашылардың, мамандардың жұмыс әрекетімен сәйкес келіп отыруы қажет.

- **Тұтынушыға жайлылық (комфорттық).** Жүйенің сервистік құралдары тұтынушының әрекеттерінің "пакеттік" және диалогтік режимдері үшін аса қолайлы болуы шарт.

Эксперименттік жүйелерді құрудың мақсаты машиналық аударманың даму жолдарын зерттеу, адамның ойлау-сөйлеу қызметін модельдеу болып табылады. Тек осы аталған тұрғыда ғана машиналық аударма мен жасанды интеллектің міндеттері бір-бірімен сәйкес келе алады. Эксперименттік жүйелер мен оларды

құрастырушы ғалымдар машиналық аударманың көшбастар рөлін атқарады да, ал олардың қол жеткізген нәтижелері өндірістік жүйелерде қолданыс табуы мүмкін. Мысалы, Саарбрюкен университетінің SUSY атты эксперименттік жүйесін құру арқылы алынған нәтижелері бойынша MARIS атты өндірістік жүйе құрылуда және SUSANNAH атты аудармашының автоматтанған жұмыс орны зерттеліп, жасалуда.

2. **Жүйенің автоматтану дәрежесі.** Практикалық тұрғыдан қарастырғанда, машиналық аударма жүйелерінің барлығы да адамның (маманның) араласуын қажет етеді. Бірақ бұл өлшемді қолдану жүйені екі қарама-қарсы көзқараста топтауға мүмкіндік береді: редакциялауда қажет ететін машиналық аудармадан бастап, аудармашы-адамға көмек беретін жүйеге дейін *(аудармашының автоматтандырылған жұмыс орны жайлы Бүкілодақтық аударма жасау орталығы құрастырған СПАС атты жүйені қара).*

Шынында да, бұл жүйелердің негізгі айырмашылығы адамды аударма жұмысына тарту стратегиясы – жүйе арқылы алынған нәтижені редакциялау немесе машинасыз (дәстүрлі) аудару кезінде машиналық аударма жүйесінің аса күшті сөздіктер мүмкіндігін пайдалану. Осыған қарай редакциялау кезеңі мен оның көлемі айырым табады. Мысалы, машиналық аударманың өндірістік жүйелерінде аударылған материалдың 20 пайызын қамтитын бөлігі үшін соңынан редакциялау (постредактирование) тәсілі кең түрде қолданылады *(ALPS, ENGSPAN, LOGOS, METAL, SPANAM жүйелерін салыстыр).* Әсіресе, бұл жайт мәтіндер жиынын пакеттік режимде өңдеуді қарастыратын жүйелерге тән болып келеді.

Машиналық аударма нәтижесі, ең алдымен, әрі қарайғы өңдеулерді қажет ететін бастапқы аударма нәтиже болатындығын ескере отырып, машиналық аударма мәселесін зерттеуші ғалымдар арнайы сервистік жүйені құрудың және практикалық машиналық аударма жүйесінің нәтижелігін бағалау үшін ондай сервистердің деңгейлік дәрежесінде де ескеру қажет екендігін байқауда. Мұндай сервис, әдетте, редакциялау құралдарын, сөздік жүйелерін интерактивті түрде енгізу құралдарын (автоматтанған лексикондарды) және т.б. өз құрамында сақтайды.

Ақпарат ағымының жаңа толқыны машиналық аударма жүйелерін күнделікті өмірде кең түрде қолдану қажеттігін айқындап отыр, сондықтан қажетті ізденістер барысында жоғарыда аталған мәселелер толық түрде ескерілуі керек.

Автоматический перевод. 1949–1963 гг.: Критико-библиографический справочник / Сост. И.А. Мельчук, Р.Д. Равич. М., 1967. 516 с.

Автоматический перевод: Сб. статей / Под ред. и предисл. О.С.Кулагиной, И.А.Мельчука. М., 1971. 357 с.

Андреев Н.Д. Машинный перевод и проблема языка-посредника // Вопр. языкозн. 1957. № 5. С. 117-121.

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистическое обеспечение системы "ЭТАП - 2". М., 1989. 295 с.

Бар-Хиллел И. Будущее машинного перевода // Науч. докл. высшей школы. Филол. науки. 1962. № 4. С. 203-206.

Белопольская А.Р., Ковригин А.Б. Исходный этап независимого анализа немецкого простого распространенного предложения // Материалы по матема-тической лингвистике и машинному переводу. Вып. 2. Л., 1963. С. 150-160.

Беляева Л.Н. Применение ЭВМ в лингвистических исследованиях и лингводидактике. Л., 1986. 84 с.

Беляева Л.Н., Откупщикова М.И.. Автоматический (машинный) перевод // Прикладное языкознание: Учебник / Отв. ред. А.С. Герд. СПб.: Изд-во С. – Петербург. ун-та, 1996. 528 с.

Воронин В.А. Система независимого конфигурационного анализа китайского текста при машинном переводе // Машинный перевод. М., 1961. С. 135-172.

Дрейзин Ф.А., Рашитов Р.С. Принцип синтаксического анализа татарской фразы // Машинный перевод. М., 1961. С. 295-303.

Кремнева Н.Д., Морева В.Ю., Ульянова Г.С. Принципы построения грамматики русского синтаксического анализа с учетом предпочтения и порядка замыкания связей // Лингвистические проблемы функционального моделирования речевой деятельности. Вып. 3. Л., 1976. С. 139-144.

Кулагина О.С. Исследования по машинному переводу. М.,1979. 319 с.

Кулагина О.С., Мельчук И.А. Машинный перевод с французского языка на русский // Вопр. языкозн. 1956. № 5. С. 111-121.

Лейкина Б.М., Никитина Т.Н., Откупщикова М.И. и др. Система автоматического перевода, разрабатываемая в группе математической лингвистики ВЦ ЛГУ // Науч.-тех. информация. 1996. № 1. С. 40-50.

Леонтьева Н.Н. Система французско-русского перевода (ФРАП): лингвистические решения, состав, реализация // Машинный перевод и прикладная лингвистика: Проблемы создания системы автоматического перевода. Вып. 271. М., 1987. С. 6-26.

Марчук Ю.Н. Проблемы машинного перевода. М., 1983; Методы моделирование перевода. М., 1985. 199 с.

Машинный перевод: Сб. статей. М., 1957. 314 с.

Мельчук И. А. Работы по машинному переводу в СССР // Вестн. АН СССР, 1959. № 2. С. 43-47; Опыт теории лингвистических моделей "смысл – текст". М., 1974. 314 с.

Молошная Т. Н. Алгоритм машинного перевода с английского языка на русский // Проблемы кибернетики. Вып 3. М., 1960. С. 209-272.

Откупщикова М. И. Синтаксис связного текста. Л., 1982. 103 с.

Панов Д. Ю. Автоматический перевод. 2-е изд. М., 1958. 72 с.

Переводная машина в Таллине в 1924 г. // Сообщения по машинному переводу. Сб. 1. Таллин, 1962. С. 95-96.

Переводная машина П. П. Траянского: Сб. матер. о переводной машине для перевода с одного языка на другие, предложенной П.П.Троянским в1933 г. / Отв. ред. Д. Ю. Панов. М., 1959. 52 с.

Пиотровский Р. Г. Инженерная лингвистика и теория языка. Л., 1979. 112 с.

Синтаксический компонент в системах машинного перевода // ВЦП. Сер. 2. Машинный перевод автоматизация информационных процессов: Обзорная информация. Вып. 5. М., 1981. 130 с.

Шаляпина З. М. Англо-русский многоаспектный автоматический словарь (АРМАС) // Машинный перевод и прикладная лингвистика. Вып. 17. М., 1974. С. 7-67.

Шумилина А. Л. Вопросы анализа личных местоимений третьего лица // Лингвистические исследования по машинному переводу. Вып. 2. М., 1961. С. 132-140.

Цейтин Г. С., Откупщикова М. И., Лейкина Б. М. и др. Система анализа текста с процедурным представлением словарной информации // Актуальные вопросы практической реализации систем автоматического перевода. Ч. 1. М., 1982. С. 205-231.

Чижаковский В.А., Беляева Л.Н. Тезаурус в систамах автоматической переработки текста. Кишинев, 1983. 163 с.

Hutchins W. J. Transiation: Past, Present, Future. Harwood, 1986.

Machine Transiation: Theoretical and Methodoiogikai Issues // Ed. by S. Nirenburg. New York, 1987. 350 p.

БАСПА ІСІН АВТОМАТТАНДЫРУДЫҢ ЛИНГВИСТИКАЛЫҚ НЕГІЗДЕРІ

Баспа ісін автоматтандыру тек авторлық шығарманың түпнұсқасын полиграфиялық түрге келтіруді электронды есептеу машинаға (ЭЕМ-ға) толығымен жүктеп қана қоймайды, сонымен бірге бірқатар лингвистикалық өңдеу әрекеттерін автоматтандыру мәселесімен де айналысады. Баспа ісіндегі мәтінді өңдеуге қатысты лингвистикалық жабдықтау кезеңдерінің сипатын қысқаша түрде баяндайық.

Баспа ісін автоматтандырудың кезеңдері. Қазіргі кезде мұндай әрекеттің мынадай кезеңдерін бөліп қарастыруға болады:

1) **авторлық кезең** (мәтін авторының шығармасын ақпарат сақтайтын тетікке көшіруі);

2) **редакциялық кезең:**

а) авторлық түпнұсқа мәтінге корректура (түзету) жүргізу мен редакциялау;

ә) редакциялық баспа қызметі (рецензиялау, корреспонденцияларды өңдеу, деректер базасынан ақпарат іздеу және т.б.);

3) **құрылымдау кезеңі** (мәтіннің әр элементіне полиграфиялық параметрлерді белгілеу: терме мен бет пішімі, шрифтердің кегелі мен гарнитурасы және т.б.);

4) **полиграфиялық қайта туындату (воспроизводство) кезеңі** (мәтіннің әр элементі үшін белгіленген полиграфиялық параметрлерді фототерім машиналары арқылы сыннан өткізу);

5) **таралымды белгілеу кезеңі,** яғни мәтінді ұсынылған дана санына сәйкес басып шығару (егер басылым дәстүрлі қағаз түрінде болса);

6) **тарату кезеңі** (басылымды ақпарат сақтайтын тетік арқылы немесе байланыстың электрондық торабы арқылы баспа қорынан оқырманға ұсыну).

5-кезеңде тілдік деректер толығымен қатыспайды, ал басқа кезеңдерде оның тек орташа дәрежесінде ғана қатысы бар. Тілдік деректердің ең күшті қажеттілігі — 2-кезең.

Баспа ісін автоматтандыру кезеңдері хронологиялық жағынан бір-біріне тұтасып, бірігіп кетуі мүмкін. Мысалы, шығарма авторы мәтінді өзі теріп, алғашқы түзетулерді өзі жүзеге асыруы да мүмкін. Сондықтан лингвистикалық жабдықтау мәселесі баспа ісін автоматтандырудың әр кезеңінде де кездеседі.

Енді жоғарыда аталған кезеңдердің әрқайсысына байланысты лингвистикалық жабдықтаудың ерекшеліктеріне тоқталайық.

Авторлық кезең. Бұл кезеңде мәтінді теру үшін авторлар әртүрлі мәтіндік процессорларды қолданады, яғни мәтінді теру мен оған өзгертулер жүргізуде компьютерлік программаларды пайдаланады. Мұндай мәтіндік процессорларда тілдің мынадай қызметтері орын алады:

1) автордың қалаған сөзіне синонимдерді іздестіру және қажеттікке қарай, ол сөзді автоматты түрде оның синонимімен алмастыру;

2) қате жазылған сөздерді эталондық орфографиялық сөздіктен іздестіру немесе ол сөздің сөздікте жоқ екендігін айқындау;

3) тұлғалық сәйкестік бойынша эталондық сөздіктегі сөзді мәтіндегі тиісті орнына көшіру.

Аталған тілдік қызметтердің біріншісі осы авторлық кезеңде жүзеге асады, ал екіншісі мен үшіншісі келесі кезеңдердің негізін құрайды, сондықтан оларды кейінірек қарастырамыз.

Авторлық кезеңде қалам иесі қажетті деген мәтін элементтерін бөліктерге бөліп алуы үшін таңбалар жүйесінің арнайы жазылған формалданған тілін қолдануына мүмкіндігі бар. Егер автор ондай таңбаларды шығарма мәтініне енгізбей кетсе, баспа жұмысын құрылымдау кезеңінде ол міндетті техникалық редакторлардың атқарулары қажет.

Сонымен бірге бұл кезеңде автордың машиналық ақпарат сақтайтын тетікте жазылған түсіндірме және энциклопедиялық сөздіктерді автономдық сипатта қолданылуына да мүмкіндігі бар. Жүйенің программалық жабдықталуы тұтынушының өзіне қажетті деген сөздік мақаланы тез тауып алуына жағдай жасайды.

Редакциялық кезең. Бұл кезеңде баспа ісін автоматтандыру жүйесі мәтінді өңдеуге қатысты корректура жүргізу мен редакциялау қызметтерін жүзеге асырады. Мәтінді өңдеу жұмысымен қатар редакциялық-баспа қызметі де орындалады. Аталған қызметтер осындай мақсаттарға арнайы арналған редакциялық-баспа жүйесінде және баспа ісін автоматтандыруда қолданылатын басқа да жүйелерде жүзеге асады.

Баспа ісін автоматтандырудағы корректуралау мен редакциялау нысаны ретіндегі шығармалар мәтіні мынадай элементтерден тұрады:

- жәй ғана мәтіндер (графемалар, яғни әріптер, морфемалар, сөздер, сөз тіркестері, сөйлемдер, фразадан тыс тұтастық; топтамалар, яғни бөлімдер, тараулар және т.т.; және дискурс, яғни бүтін түріндегі мәтін);
- күрделі мәтіндер, яғни математикалық өрнектер мен кестелер;
- шығарма аппаратын құрастырушы мәтіндер (титулдық бет, басылым мәліметтері, сілтемелер және т.б.);
- қызметтік мәтіндер, яғни мәтін элементтерінің таңбалары және полиграфиялық безендіруге қатысты бұйрықтар.

Редакциялау кезеңінде баспа ісін автоматтандыру әрекетіне қатысты авторлық мәтінді өңдеудің мынадай екі түрлі амалын атауға болады:

1) мәтіндерді түпнұсқаға сәйкестендіру;

2) мәтіндерді қазіргі кездегі тілдік нормаға сәйкестендіру.

Баспа ісін автоматтандыру ісінде бірінші процесті *корректура* жүргізу деп, ал екіншіні – *редакциялау* деп атайды. Корректура жасау мен редакциялау істерінде де екі түрлі әрекеттерді ажыратуға болады:

а) бақылау амалы, яғни ауытқуларды табу;

ә) қайтақұру (реконструкция) амалы, яғни бақылау барысында анықталған ауытқуларды түзету.

Бұл айтылғаннан мынадай қорытынды шығаруға болады: бақылау амалының нысаны ретінде барлық мәтін көлемі алынады да, ал қайтақұру нысаны ретінде – ауытқулар түрлері орын алатын мәтін бөліктері алынады.

Енді ауытқулардың түрлеріне жататындар:

а) бұрмалаулар (түпнұсқадан айырым табатын ауытқулар);

ә) қателер (нормадан ауытқу);

б) қате жіберу мүмкіндіктері (түпнұсқадан да және нормадан да бірдей кететін ауытқулар).

Әрине, редакциялау барысында бақылауды қажет ететін (баспаның, логикалық, техникалық және т.б.) норма мәселесінің барлығы бірдей емес, тек лингвистикалық және психолингвистикалық түрлері ғана орын алады. Бұл топтардың ішінен мынадай нормалық топ түрлерін бөліп қарастыруға болады:

а) тізімдер (мысалы, неше түрлі сөздіктер, қысқартулар тізімі немесе өлшем бірліктерінің тізімі);

ә) үлгілер (мысалы, сөйлемдер мен қайталаным кезіндегі тыныс белгілерді қою үлгісі);

б) құрылымдар немесе модельдер (мысалы, басылымның құрылымы немесе сөйлем моделі);

в) параметрлер (мысалы, бір бетте қатарынан төрт тасымалдың бірдей кездеспеуі немесе аннотация көлемінің 600 таңбадан аспауы).

Корректурадан немесе редакциялаудан өткен автор мәтіннің сапасын анықтау үшін мынадай шарттардың орындалғаны жөн:

- корректура үшін – мәтін көшірмесінің бұрмалану дәрежесі мына өрнекпен анықталады:

$$E=e/M,$$

мұндағы e – мәтінде кездесетін бұрмалану саны, M – авторлық баспатабақ бойынша есептелген мәтін көлемі;

- редакциялау үшін – қателесу дәрежесі:

$$S=r/M,$$

мұндағы r – мәтіндегі қате саны.

Машиналық (компьютерлік) корректура жүргізу мен редакциялаудың кейбір технологиялық ерекшеліктеріне тоқталайық. Мұндай ерекшеліктердің ең алғашқысы ретінде мәтінді өңдеудің үш кезеңін атап айтуға болады.

Бірінші технологиялық ерекшелікке жататындар:

- корректураның алдын алатын әрекет (предкорректура) және редакциялаудың алдын алатын әрекет (предредактирование).

Мысалы, мәтін элементтерінің әрбіреуіне сәйкестікке қойылатын таңба;

- интеркорректура (интерредакциялау), яғни компьютер көмегімен орындалатын бақылау мен қайтақұру әрекеттері;

- корректурадан кейінгі әрекет (посткорректура) және редакциялаудан кейінгі әрекет (постредактирование), яғни техникалық құралдар арқылы орындауға келмейтін жағдайдағы корректор мен редактордың амал-әрекеттері.

Екінші технологиялық ерекшелікке корректура мен редакциялаудың әр дәрежедегі автоматтану сипаты жатады:

- компьютерленген корректура және компьютерленген редакциялау (мәтіндік процессор арқылы компьютер жадына жазылған мәтінге адамның араласуы арқылы ғана жүзеге асатын бақылау мен қайтақұру әрекеттері);

- автоматтанған корректура мен автоматтанған редакциялау (мәтінді бақылау жұмысы ЭЕМ арқылы жүзеге асады да, ал қайтақұру әрекеті – корректор немесе редактор арқылы);

- автоматтанған корректор мен автоматтанған редакциялау (мәтінді бақылау мен қайтақұру жұмыстары ЭЕМ арқылы орындалады, ал адам қалыпты жағдаят бұзылса ғана араласады).

Мәтінді бақылау немесе қайтақұру және де корректура немесе редакциялау әдістерінің тиімділігін анықтау үшін мына өрнекті қолдану қажет:

$$E=(O_1-O_2)/O_1,$$

мұндағы E – мәтінді өңдеудің тиімділігі, O_1 – мәтінді өңдеуден бұрынғы ауытқулардың саны, O_2 – өңдеуден өткеннен кейінгі мәтіндегі ауытқулар саны.

Тәжірибеде дәлелденіп және ғылыми зерттеулердің нәтижелері бойынша расталып отыратын жайт – мәтін ағымын өңдеу барысында бұрмалану мен қателерді толығымен жоюдың мүмкін еместігі. Сондықтан, ең дұрысы, өңделу жұмысының дәрежесі жайлы ғана сөз ету (бұрмалану дәрежесі немесе қате кету дәрежесі). Осындай жағдайды ескере отырып, бұрмалану дәрежесі мен мәтін түріне қарай әдебиеттерде тиісті талаптар қойылады.

Осы айтылғандарға байланысты ең жиі қолданылатын машиналық корректура әдістерін атап өтейік. Оларды машиналық қайтақұру және машиналық бақылау әдістері деп екіге ажыратуға болады. Машиналық бақылау әдістері, іштей тағы да тұпнұсқаны қолдануға негізделген және белгіленген нормаға сай келетін символдар тізбегіне негізделген деп екіге ажыратылады.

Шығарманың тұпнұсқа мәтінін пайдалануға негізделген әдістер ішінен ең кең қолданыс табатыны – қайталау (дублирование) әдісі. Оның мәні мынада. Мәтінді ақпаратты теру құралы арқылы таспаға (машина тетігіне) екі рет түсіріп алып, олардағы әрбір символды компьютерлік программалар көмегімен бір-бірімен салыстыру қажет. Егер қайталанған мәтіндер ішінде бір ғана позиция бойынша екі түрлі символ кездессе, онда қайталанған мәтіндердің біреуі бұрмаланған, яғни сол позицияда қате бар деген сөз. Мұндай бақылау әдісінің нәтижелілігі маман-корректордың іс-әрекетімен салыстырғанда жоғары және ол – 0,96-ға тең.

Нормаланған символдар тізбегіне негізделген бақылау әдістерінің ішінен ең көп тарағаны – сөздік әдісі. Оның мәні мынада: қайсыбір тілдің орфографиялық сөздігі ЭЕМ-ның есте сақтайтын сыртқы тетігіне жазылады да, түзетуге қажетті мәтіннің әрбір сөзі эталон ретінде алынып отырған ЭЕМ жадындағы сөздік бірліктерімен сәйкестендіріледі. Егер іздеген сөз сөздіктен табылса, онда ол сөз дұрыс жазылды деп саналады да, ал сөздік бойынан табылмаса – потенциалды дәрежеде (барынша) бұрмаланған сөз деп есептеледі. Мұндай бақылау әдісінің нәтижелігі қайталау әдісіне қарағанда төмен дәрежеде және ол 0,5 пен 0,7 аралығындағы шамаларға тең.

Бұл жерде мынадай сұрақ туындауы мүмкін: норма бойынша (орфографиялық сөздік арқылы) мәтін көшірмесінің тұпнұсқадан ауытқу жағдайын қалайша бақылауға болады? Мұндай мүмкіндіктің орын алуы, көпшілік жағдайда, көшірме мәтінінің тұпнұсқа мәтінінен ауытқуы нормадан да ауытқу болып саналуына байланысты. Қазіргі кездегі баспа ісінің тәжірибесі бойынша, қайтақұру (реконструкция) әрекеті әр уақытта автоматты түрде жүзеге асады деп саналады. Яғни мәтіндегі бұрмалауларды автоматты түрде іздеп тауып, оларды адам қолымен түзетуге болады деп ұйғарым жасалады. Солай бола тұра, формалданған мәтіндер үшін қайтақұру әрекетінің нәтижелілігін 0,8 бен 0,9 аралығында жүзеге асыруға болатын әдістер де бар.

Компьютер арқылы редакциялауда бақылау нормаларының саны редакциялау жүйесінің "редакциялау қуаттылығын" білдіреді. Бұл қуаттылық, айтарлықтай дәрежеде, шығарма мәтіні элементтерінің осы мақсатқа қаншалықты сай таңбалануына да байланысты болады.

Лингвистикалық нормалар, негізінен, жәй (қарапайым) мәтіндерді тексеруге мүмкіндік жасайды. Мұндай бақылау әрбір сөзге әртүрлі ақпараттар

(морфологиялық, синтаксистік, прагматикалық және т.б.) қоса тіркеліп берілген сөздікке негізделеді. Лингвистикалық әдістердің ең қарапайым түрлері сөздердің кейбір критерийлері (өлшемдері) бойынша таңдалып алынған қайсыбір жиынды пайдалануға немесе, керісінше, кейбір сөздерді пайдаланбауға мүмкіндік тудырады. Күрделіленген әдістер сөздер арасындағы грамматикалық байланыстың, тыныс белгілерінің, мәтінді абзацтарға бөлудің, қысқартылған сөздердің және сөз тіркестерінің берілулерінің және т.б. дұрыстығын бақылауды жүзеге асырады.

Психо-лингвистикалық нормалардың да барынша кең спектрі қолданыс табуда (*Мучник, 1985).* Бұл нормалар белгілі оқырман қауымына арналады және сөз, сөйлем немесе фразадан тыс бірліктердің мүмкін болатын максимум "ұзындығын" белгілейді (бұларды ***тез оқырлық*** сипаттамалар деп те атайды). Осымен бірге, кейбір әдебиеттердің түрлеріне қарай, арнайы нормалар түрлері де белгіленеді. Мысалы, шет тілдерінің оқулықтары үшін мына жайттар тексеріледі:

- мәтін жаңалығы, яғни мәтіндегі оқушыға таныс емес (жаңа) сөздер санының мәтіндегі барлық сөздер санына қатынасы (нормалық мәні – 0,015 пен 0,50 аралығында);
- мәтіннің адаптациялануы (бейімделуі), яғни мәтіндегі бұрыннан да таныс сөздер санының оның жалпы сөздік қорына қатынасы (нормалық мәні – 0,237 мен 0,253 аралығында);
- баяндаудың қиындық сипаты (напряжение изложения), яғни оқушыға таныс емес сөздер санының оның жаңадан меңгерген сөздерінің санына қатынасы.

Лингвистикалық жабдықтау жүйесін редакциялық-баспа қызметін автоматтандыруға да қолдануға болады. Мәселен, мұндай мақсат үшін әртүрлі арнайы ақпаратты-іздестіру тілдері құрастырылады және олар арқылы баспа мекемесіне жолданған хаттар (корреспонденциялар) өңделеді.

Құрастыру кезеңі. Бұл кезеңде басылымды құрастыру мақсатында тілдердің екі түрі қолданылады:

а) полиграфиялық безендірудің бұйрықтар тілі;

ә) мәтін элементтерін бөліп алуға қажетті таңбалар тілі (язык меток).

Бұл тілдер әртүрлі ұстанымдарға негізделіп жасалады. Жүйелерде таңбалар мен бұйрықтар саны, көбінде, 20 мен 100 сандарының аралығында орналасқан.

Көпшілік жағдайда полиграфиялық бұйрықтар мынадай құрылымға ие болып келеді:

а) бұйрық белгісі;

ә) бұйрық түрі;

б) бұйрық параметрі (өлшемі).

Мәтін элементтері таңбасының (меткасының) құрылымы, негізінен, полиграфиялық бұйрық құрылымымен бірдей келеді. Бірқатар жүйелерде мәтін элементтеріне таңбалар тілі (язык меток) қолданылмайды да, ал оның орнына мәтін элементінің алдына полиграфиялық безендіру бұйрығы қойылады. Бұйрықтар мен таңбаларды (меткаларды) мәтінді полиграфиялық қайта туындату кезеңінен бұрынғы кезеңдердің барлығында да бірдей қоя беруге болады.

Полиграфиялық қайта туындату кезеңі. Бұл кезеңде баспа ісін автоматтандырудың лингвистикалық жабдықтау мүмкіндіктерінің ішінен, берілген форматтағы жолды қоспау үшін, тек бір ғана функция пайдаланылады. Ол функция сөзді тасымалдау қажеттігі үшін әрекет ететін компьютерлік программа түрінде көрініс табады. Мұндай программалар полиграфиялық безендіру бұйрықтарын өтеу үшін арналған мәтіндерді өңдеудің автоматтанған жүйелерінде қолданыс тауып жүр.

Сөздерді тасымалдау программалары баспа ісін автоматтандырудың алдыңғы кезеңдерінде қызмет ететін басқа да жүйелерде қолданылуы мүмкін. Мысалы, мәтіндік процессорлар мен редакциялық-баспа жүйелерінде қолданылады. Бірақ бұл жүйелерде сөзді тасымалдау программасына қойылатын талап полиграфиялық қайта туындату кезеңіне қарағанда айтарлықтай қатал емес.

Сөзді тасымалдау алгоритмін екі үлкен топқа бөліп қарастыруға болады:

- сөздіктер арқылы, яғни бұл жағдайда компьютер жадында қайсыбір тілдің тасымалданған сөздерінің эталондық нұсқасы сақталады. Мұндай сөздердің көлемі 20-40 мың және бірнеше тілдер түрі (көбінде, оннан аспайтын тілдер) қамтылады;
- сөздіксіз, яғни бұл жағдайда тасымалдың белгісі тасымалдау ережесі бойынша компьютерлік программа көмегімен жүзеге асады.

Әрине, сөздіктік әдіс бойынша тасымалдау ісі аса нәтижелі болғанымен, мұнда әрекет ету мезгілі мен компьютер жадын үнемдеуден көптеп ұтылады. Ал сөздіксіз әдіс уақыт пен жадыға байланысты тиімді болғанымен, тасымалдау әрекетінің сапасы өте төмен.

Сөздіксіз әдістер ішінен мынадай кішігірім топтарды бөліп қарастыруға болады:

- морфемдік әдіс арқылы, яғни тасымалдау белгісін сөздің морфемдік құрамына қарай қою;
- морфемсіз әдіс арқылы, яғни тасымалдау белгісін буындар шекарасына қарай немесе құрамында дауысты және дауыссыз әріптердің екіәріптік тіркестерінің (биграммалардың) болуына қарай қою.
- қисындастыру әдісі. Бұл әдіс бойынша тасымалдау белгісі сөзформа құрамында префикстердің, суффикстердің және буындардың немесе биграммалардың болуына қарай қойылады.

Сөздіксіз әдістер арқылы жүргізілген зерттеулер нәтижесі бойынша, бұрыс қойылған тасымалдар саны, кері функционалдық байланыстың әсерінен қойылмай кеткен (жоғалған) тасымалдар санына қатысты болып келеді. Мысалы, егер қайсыбір алгоритм бойынша бұрыс қойылған тасымалдардың саны 1,12 пайызға тең болса, онда қойылмай кеткен (жоғалған) тасымалдардың саны 48,41 пайыз құрайды. Сондай-ақ, керісінше, басқа бір алгоритм бойынша қойылмай кеткен (жоғалған) тасымалдардың саны 7,40 пайыз болса, бұрыс тасымалдардың саны 11,70 пайызға дейін өседі (*Будаар және т.б., 1985).*

Тарату кезеңі. Баспа ісін автоматтандырудың бұл кезеңі бойынша арнайы тасымалдау (немесе байланыс торабы) құралдары арқылы даяр машиналық ақпарат, яғни басылымдар нәтижесі тұтынушылар қауымына таратылады. Басылымдар – *мерзімсіз* немесе *мерзімді басылымдар* болып екіге ажыратылады. Мысалы, мерзімді басылымдарға "электронды" түрдегі журналдар мен газеттерді жатқызуға болады. Мұндай "электронды" басылымдар түрлері бірқатар Батыс Еуропа, Солтүстік Америка елдерінде және Жапония мен тағы да басқа елдерде дәстүрге айналып отыр.

Баспа ісін автоматтандырудың бұл кезеңінде мерзімді басылымдардың әр бөлігінен тұтас бір басылым құрастыру үшін, негізінен, лингвистикалық жабдықтау мүмкіндіктерінің әртүрлі рубрикаторлардың иерархиялық бір деңгейінде техникалық себептерге байланысты тек 10 рубрика ғана кездесіп жүр.

Баспа ісін автоматтандыруды лингвистикалық жабдықтау жүйесінің құрылымы. Мұндай құрылым *автономды* және *жүйелі* деп екіге бөлінеді. Лингвистикалық жабдықтаудың автономды түрінде түсіндірме және энциклопедиялық сөздіктер қамтылады. Жүйелі түрдегі лингвистикалық жабдықтауда деректердің тірек базасы, нормалардың жиыны және компьютерлік программалар сақталады.

Деректердің тірек базасының құрамы мынадай ақпараттардан тұрады: антонимдік, синонимдік, орфографиялық сөздіктерден, морфологиялық, синтаксистік, семантикалық және прагматикалық ақпараттар сөздіктерінен (жалпытілдік және салалық), эталондық тасымал сөздіктерінен (әртүрлі тілдер үшін), морфемалар тізімінен, өлшем бірліктері қысқартуларының тізімінен, персоналий тізімінен, полиграфиялық бұйрықтар тізімінен, мәтін элементтері таңбаларының тізімінен, библиографиялық сипаттаудың қалыпты элементтерінің тізімінен және арнаулы ақпараттық-іздестіру тілдері рубрикаторларының тізімдерінен.

Нормалар жиыны іштей мынадай лингвистикалық нормалар жиындарының деңгейі бойынша бөлінеді: графемдер, морфемдер, фразадан тыс бірліктер, блоктар (топтамалар), дискурстер; жоғары дайындық деңгейіндегі оқырмандар үшін психолингвистикалық нормалар жиыны, төменгі дайындық деңгейіндегілерге – тасымалдау ережелерінің жиыны (әртүрлі тілдер үшін).

Компьютерлік программалар топтамасы мынадай құрамнан тұрады: авторлық және редакциялық кезеңдеріне арналған программалардан, құрастыру мен полиграфиялық қайта туындату кезеңдері үшін жазылған программалардан және тарату кезеңіне арнайы жазылған программалардан.

Лингвистикалық жабдықтаудың құрастырушылары бөлек және бірге көрініс табады. Мысалы, лингвистикалық ақпаратты сақтайтын сөздік құрамында орфографиялық және синонимдік сөздіктер де бірге берілуі мүмкін. Міне, осыған қарамай, лингвистикалық жабдықтаудың әрбір компонентін жеке бір нысан ретінде қарастыруға болады.

Соңғы жылдары дамыған дербес компьютерлер мен лазерлік принтерлердің негізінде ***үстелүстілік баспа жүйесі*** деп аталатын жүйелер құрыла бастады. Желі режимі бойынша да әрекет ете алатын мұндай үстелүстілік баспа жүйелер өзіне жүктелген қызметті қарастырылған кезеңдердің барлығында бірдей (тарату кезеңін есепке алмағанда) атқара алады.

Сонымен, лингвистикалық жабдықтау жүйесіне қойылатын негізгі талап – тек бір ғана лингвистикалық сөздік бойынша әрі синоним сөздерді іздестіруді, әрі корректураны, әрі редакциялауды, әрі сөздерді тасымалдауды жүзеге асыру.

Міне, осы аталған мәселелерді компьютерлік лингвистика тұрғысынан зерттеу баспа ісін автоматтандыру әрекетін жаңа сипаттағы жоғары деңгейге көтерері сөзсіз.

Будаар Я. И., Карская С. Д., Партыко З. В. Сравнительный анализ алгоритмов переносов слов русского языка // Научно-техническая информация. Сер. 2. 1985. № 12. С. 20-30.

Мартин Дж. Видеотекст и информационное обслуживание общества. М., 1987. 184 с.

Мучник В. С. Человек и текст: Основы культуры письменной речи. М., 1985. 256 с.

Партыко З. В. 1) Методы машинной корректуры и машинного редактирования. М., 1983. 41 с. (Изд. дело: Обзорн. информ. / НИЦ "Информпечать". Вып.5); 2) Статистика ошибок при корректуре и редактировании текстов. М., 1989. 56 с. (Изд. дело: Обзорн. информ. / НИЦ "Информпечать". Вып.3); Лингвистическое обеспечение автоматизированного издательского процесса / Прикладное языказна-ние: Учебник, отв. ред. А.С. Герд. СПб: Изд-во С.-Петербург. С. 399-407.

ҚАЗАҚ ТІЛІНІҢ КОМПЬЮТЕРЛІК ҚОРЫН (БАЗАСЫН) ҚҰРАСТЫРУ ЖАЙЛЫ

Кеңес Одағы кезіндегі кейбір ғажайып бастамалар заман өзгерісіне қарай құбылып, шешімі кезек күттірмейтін көптеген мәселелерге көзқарас басқаша қалыптасып, әрі қарайғы жалғасы табылмай қалғаны қазіргі кезде құпия бола қоймас. Солардың бірі – түркі тілдерінің, оның ішінде қазақ тілінің машиналық (компьютерлік) қорын жасау мәселесі.

Осы айтылған мәселенің қысқаша тарихы мен мән-жайына тоқталайық.

Электронды есептеу машиналарының (ЭЕМ) көмегімен ұлттық тілдерді зерттеу мен одан туындайтын ақпараттарды автоматты түрде танып-білудің негізінде XX ғасырдың 80-жылдарынан бастап тілдердің машиналық қорын (базасын) жасау жұмысы жайлы ой-пікірлер жүзеге аса бастаған болатын. Осыған байланысты 1983 жылы Мәскеуде шақырылған Бүкілодақтық ғылыми конференцияның күн тәртібіне орыс тілінің машиналық қорын жасау проблемасы алғаш рет қойылды. Бұл ғылыми мәжілісте ЭЕМ-нің жадына түсірілетін тілдік қордың құрамы мен құрылымы қандай болуы тиіс екендігі, бірінші кезекте мұндай қор қандай міндеттер атқаратыны, тіл көздерінің ауқымы қандай болмақтығы жайлы және т.б. көптеген мәселелер қаралған болатын. Сонымен бірге электрондық тіл қорын дәйекті түрде тұтыну үшін қандай ұйымдық және техникалық шаралардың атқарылу қажеттігі де ортаға салынып, жан-жақты талқыланды.

Аталған ғылыми конференцияның шешімдерінің орындалу жағдайын қадағалау мен жаңадан туындайтын мәселелерді шешу қажеттігінен 1987 жылы Мәскеуде ашылған Бүкілодақтық екінші ғылыми конференцияда осы саладағы ғалымдар тағы да бас қосты. Бұл ғылыми форумға Қазақстанның және басқа да ұлттық республикалардың ғалымдары қатысты.

Мәжілісте өткен төрт жыл аралығында орындалған жұмыстардың нәтижелерінің сөз болуымен қатар, басқа да ұлт тілдерінің машиналық қорларын жасау мүмкіншіліктеріне байланысты соны мәселелер қарастырылды. Ерекше атап кетерлік жәйт – осы ғылыми жиында орыс, украин, грузин, эстон тілдеріне қатысты баяндамалардың қатарында қазақ тілінің машиналық қорын жасау мәселесі де өз алдына тыңдалды.

Сөйтіп, орыс тілінің машиналық қорын жасау мәселесіне арналған екінші ғылыми конференцияның шешімдерінде қазақ мәтіндерін ЭЕМ жадына енгізу және оны автоматты жолмен статистикалық тәсілдермен зерттеу бағытындағы қазақ ғалымдарының қол жеткізген нәтижелері ескеріліп, түркі тілдерінің машиналық қорын жасайтын координациялық орталық Қазақ ССР ҒА Тіл білімі институтының құрамынан ашылсын деген дербес қаулы қабылданғаны біз үшін үлкен абырой болды.

Көп ұзамай, 1988 жылдың мамыр айында Мәскеуде Кеңес Одағы түркітанушы комитетінің XIV Пленумы ашылып, оның күн тәртібіне түркі тілдерінің машиналық қорын жасау мәселесі арнайы қойылды. Бұл ғылыми жиында негізгі баяндамашылар ретінде Одақ бойынша белгілі ғалымдар: Р.Г.Пиотровский, А.М.Щербак, В.Г.Гузев өз ой-пікірлерін ортаға салып, ондай қор жасаудың қажеттігімен қатар алда тұрған қиыншылықтарын да атап өтті (*Гузев, Пиотровский, Щербак*). Ал қазақ тілінің машиналық қорын жасау мүмкіндігі жайлы мәселеге байланысты қазақстандық

ғалымдар баяндама жасады (Қ.Б.Бектаев, А.Қ.Жубанов).

Тілдің машиналық қорының қажеттілігін алғаш дәлелдеп берген ғалым – информатика саласының белгілі маманы А.П.Ершов тілдің машиналық қорын құрастыру мәселесіне үлкен мән берген болатын (*Ершов, 1979*). Оған ғалымның 1978 жылы айтқан мына сөздері айғақ: "Орыс тілін сүйетін қоғамның (формалды қоғам) танымынша, орыс тіл білімі дегеніміз әзірге ұйықтап жатқан Гулливер тәрізді, сондықтан менің ойымша, оның оянатын және өзі туралы бар дауыспен жар салатын кезі енді жетті. Ал ЭЕМ бұл ретте таптырмайтын қоңыраулы сағат секілді" (*Машинный фонд..., 7-12-б.*). Бұл айтылғандар қазақ тілінің қазіргі жағдайының компьютерлік лингвистика саласына қатысты әлі де өз күшін жоғалтпаған тәрізді. Ал А.П.Ершов сол кездің өзінде-ақ аса білгірлікпен болжап, болашақ жасалатын орыс тілінің машиналық қорын тіл жөніндегі білімдердің аса қуатты концентраты (қоспасы) болады деп сенген.

Мұндай формалды компьютерлік жүйе тілдің жаратылысына сәйкес әрі тең көлемді болуы қажет, сөйте тұра ондай тілдік модель "анатомиялық" тұрғыдан мүшеленген, ашық, бақылауға айқын, зерттеуге қолайлы әрі өзгермелі болуы тиіс деп ұйғарамыз. Алайда қазақ тілін модельдеуге қатысты біздің жетістіктерімізді үндіеуропа тіл білімінің статусымен және оның компьютерлік лингвистика бойынша алған нәтиже ауқымымен салыстырып қарайтын болсақ, біз өтпек жолымыздың әлі де бастауында тұрғанымыз айқын сезіледі.

Көптеген ғалымдардың пікірлерін қорыта келе, біз "Қазақ тілінің компьютерлік қоры" атты жүйенің түсінігі – көп тармақты әрі терең автоматтандырылған жүйені аңғартады демекпіз және оның атқаратын міндеті ақпараттық барлау мен тіл білімі саласындағы (жалпы филология бағыттағы) зерттеушілік болуы тиіс. Ал ондай компьютерлік қор жасаудың методологиялық негіздері ретінде мынадай екі тезисті басшылыққа алғанды жөн санар едік. Оның біріншісі – тіл жөніндегі деректердің қандайы болмасын лексикографиялық пішінге келтірілу қажеттігі, ал екіншісі – лексикографиялық жағынан жүйеленген тілдік мәліметтердің барлығын да компьютерлік алгоритмдер тіліне лайықтаудың керектігі.

Тілдік бірліктердің (ақпараттардың) лексикографиялық пішінде көрініс табуы олардың сөздік түріндегі жұмыстардың идеясына жақындайды, ал мұндай тұжырым тілдік ақпараттарды шоғырландыратын бірегей түйін немесе түп қазық – жеке сөз, деп ұйғарумен бірдей. Шынында да, сөз өзінің ішкі-сыртқы қызметтері жағынан және ақпарат сақтағыштық қасиеттерінің негізінде белгілі бір жүйедегі сөз қорын танытатыны сөзсіз. Сондықтан сөздің белгілі мөлшерде өзіндік сегменттері болады және олар грамматикалық, стилистикалық, фонетикалық, контекстік т.б. пішінде көрініс табады, яғни ол барлық тілдік мәліметтер жүйесін сәулелендіреді деп ұйғарған жөн.

Егер біз ана тіліміздің компьютерлік қорын құру методологиясына орыс тілінің машиналық қорын жасау тәжірибесіне сүйенсек, онда мынадай қор құрамын сақтауымыз қажет:

1) қазақ тілі компьютерлік қорының негізгі сөзтізбегі (генеральный словник);
2) қазақ мәтіндерін суреттеу қоры (иллюстративный фонд);
3) қазақ тілінің терминологиялық қоры;
4) қазақ тілінің академиялық сөздіктері мен грамматика қоры;
5) қазақ тілінің лингво-статистикалық қоры;
6) қазақ тілінің лексикографиялық қоры;
7) қазақ тілі процессорларының қоры;
8) қазақ тіліне қатысты лингвистикалық алгоритмдер мен комплексті сипаттағы теориялық және тәжірибелік мәліметтерді бірлікте қамтитын бағдарламалар қоры.

Аталған құрамдағы қазақ тілінің компьютерлік қоры тілдік мәліметтердің сала-салаға жүйеленген банк ретінде де және тұтыну қажеттігіне сай қайта жинақталған тұтас күйінде де пайдалану мүмкіндігі алдын ала ескерілуі керек.

Қазақ тілі жөніндегі деректерді жинаудың, сақтаудың, талдаудың және салыстыра зерттеудің жаңа әдістеріне көшу, сондай-ақ лингвистикалық дерек көздерінің автоматтандырылған сөздіктер мен грамматикалар тәрізді жаңа түрлерін игеру ісі өмірге жанасымды және тиімді екені сөзсіз. Әрине, бұл үшін біз жалпыфилологиялық дәстүрлер мен мәдениетке, тілді танып білуге, практикалық жаңа міндеттер тәжірибелерін мұқият ескеруге тиіспіз.

Болашақ қазақ тілінің компьютерлік қорының құрамындағы "қазақ мәтіндерін суреттеу қорының" мақсаттары мен міндеттері қандай болмақ?

Біздің ұйғаруымызша, мұндай жүйелерде тұтас қалыптағы мәтіндер сақталуы қажет те және тұтынушының сұранысы бойынша сол мәтіндердің үзінділері, қор ішінде тұрған олардың нақты орны, сөзтұлғалардың статистикасы, сөзнұсқағыштардың толық тізімдері, сөздерге телулі цитаталар, жиілікті және кері әліпбилі сөздіктер көрініс табуы қажет.

Компьютерлік қордың лексикалық сипаты түпнұсқа мәтіндерінің типіне байланысты деп санай отыра, қазақ тілінің жүйесін танытатын сөйленіс жағдаятына: көркем әдебиет тілін, газет-журнал тілін, ғылым мен техника тілін, ұйымдастырунұсқау құжаттары тілін, заң орындары мен сот мекемелер тілін, оқу орындары тілін, сауда, дәріхана, емханалардағы және т.б. сөйленіс түрлерін жатқызуды қолдар едік. Әрине, қордың бұл түріндегі компьютер жадына орналастыруға қиындық туғызатын сөйленіс түрлеріне жататындар: диспуттар, сот процестері, жиналыстар, күнделікті әңгімелесулер, телефон арқылы сөйлесу т.т. сияқты ауызекі сөйлесу тілінің үлгілерін жазып алу мәселесі.

Келесі бір қомақты мәтіндік бірлік – ол терминдер мәселесі. Қазіргі ғылыми-техникалық және іскағаздары мәтінінің негізгі өзегі терминдер екені айқын. Жалпы сөйлеу тілінің аясындағы сөздер емес, терминдердің тілдік қоршауы (мәнмәтіні) өзгеше бір тілдік жүйе, басқаша айтқанда, терминдер өрісі болып табылады. Осы аяда өмір сүретін термин үшін барлық ғылыми шығармалардың сөз қолданыс машықтары етене, ортақ болып келеді.

Осындай маңызды жұмыстардың келесі түрі – әр дәуірдегі мәтіндердің тілдік белгілерін ажыратып ала білу мәселесі. Осы мақсатта тіліміздегі диалектілік лексиканың біркелкі еместігін ескеретін болсақ, оларды компьютерлік қорға қосу жұмыстарының қазақ диалектологтары үшін ойланатын жақтары жоқ емес. Мәселен, олар бір жағынан ертедегі сөздер мен қолданыстардың ұлтқа, жазуға дейінгі үлгілері болып келсе, екінші жағынан басқа тілдерден енген сөздер, таралу өрісі шектеліп қалған сөздер немесе тіліміздегі жаңадан енген сөздер болуы да ықтимал.

Халық тілі сол халықтың жанды қазынасы болғандықтан, соның бүкіл болмысын барлық ерекшеліктерімен қоса қамту болашақ "Қазақ тілінің компьютерлік қорының" міндеті деп түсіну қажет сияқты.

Жоғарыда сөз болған қазақ тіліндегі диалектілер секілді, әрі синхрониялық, әрі диахрониялық сипатта күн кешетін тіл – фольклордың тілі және жер-су аттарының тілі немесе топонимика.

Осы аталған тіл үлгілерін жинастырып, тіліміздің компьютерлік қорын жасауда, оның бұл аталғандардан басқа да салалары мен арналары қамтылары сөзсіз.

Формалды пішіндегі қазақ тілінің компьютерлік қорын жасау мақсатымен байланысты күтілетін ең маңызды мәселелер мыналар деп білеміз:

а) тұтынушыға қажет болған белгілі бір сөзді қандай тіл көздерінен, мәтіндерден, тіл айырымдарынан іздестіру қажет;

ә) іздеуге тиісті сөз қарастыратын мәтінде қамтылған ба, жоқ па?

б) графикалық және фонетикалық пішіндерін қоса есептегенде сөздің мүмкін болған барлық түрлері мен тұлғасы айқындалып болды ма?

в) белгілі бір сөзге қатысты мәтін үзінділерінің (мәнмәтіндердің) жалпы аумағы, сипаты анықталды ма?

г) белгілі бір сөз: қарастырып отырған семантикалық және грамматикалық сипатта өткен дәуірлердегі, көне ғасырдағы мәтіндерде қолданылған ба, егер қолданылса, қандай сөздіктерден көрініс тапқан?

Міне, осы аталғандар қазақ тілінің компьютерлік қорын жасау мақсатындағы күн тәртібіне қойылатын ең маңызды мәселелер болуы қажет.

Жоғарыда сөз болған Совет түркологтары комитетінің XIV Пленумына қайта оралып, мәжілісте көтерілген негізгі мәселелерге тоқтала кетейік.

Алғашында мәжіліске қатысушылар, тілдің машиналық қорының анықтамасы төңірегінде пікір алмасып, соның нәтижесінде бір түйінге келді деуге болады.

ЭЕМ-ның (компьютердің) араласуымен зерттелетін жекеленген мәселелердің нәтижелерін, мысалы, жеке тұрып лексика-семантикалық, этимологиялық, морфологиялық, тіпті, синтаксистік есептерді шешетін машиналық бағдарламалардың жиынтығын немесе ЭЕМ-ның жадына енгізілген мәтіндер мен лексикографиялық материалдардың тұтастығын сол тілдің машиналық қоры, міне осы, деп есептеуге болмайтындығы баса көрсетілді.

Мәселен, ҚР БжҒМ А.Байтұрсынұлы атындағы Тіл білімі институтында компьютерлік лингвистика саласы бойынша біршама ғылыми жұмыстар орындалды. Оларды қысқаша былайша атап өтуге болар еді:

- көркем әдебиет, көсемсөз, ғылыми-көпшілік әдебиет пен оқулықтар мәтіндерінің, алғашында ЭЕМ-ның, кейіннен компьютердің жадына көптеп енгізіліп, олар бойынша арнайы компьютерлік бағдарламалар құрастырылуының негізінде әртүрлі жиілік сөздіктердің алынуы;

- М.Әуезовтің 20 томдық шығармалар жинағының мәтіндерінің компьютер жадына енгізілуі және соның негізінде жазушы тілінің жиілік сөздіктерінің жарық көруі;

- осы тәрізді жиілік сөздіктердің негізінде қазақ сөздерінің әр стильде және әр көлемді мәтіндерде қайталануының қосынды жиілігінің, сол мәтіндерді қамту статистикасының анықталуы;

- сөзтұлғалардың құрылымдық статистикасы мен мәтіндегі пайыздық салмақтарының әр сөз таптарына қатысты айқындалуы;

- қазақ тіліндегі негізгі сөз таптарының мәтін бойында кездесулерінің математикалық статистика мен ықтималдық теория заңдылықтарына бағыну-бағынбау жағдайларының анықталуы;

- қазақ тілінің "ТІЛ – ҚАЗЫНА" атты автоматтандырылған компьютерлік картотекалық қорының іске қосылуы;

- қазақ тіліндегі мәтін мазмұнын оның тұрпатына қарай ашудың (тұрпаттаудың, формалдаудың) негізгі принциптерінің айқындалуы;

- Тіл білімі институтында жоспарлы тақырыптар бойынша түзіліп жатқан түсіндірме, екі тілдік, бір тілдік сөздіктер мәтіндерінің компьютер жадында "сөздіктер қорын" құрап, тұрақты сақталуы.

Міне, осы жұмыстар түрлерінің жиынтығын біз қазақ тілінің компьютерлік қоры деп атауымызға бола ма?

Жоқ. Себебі, бұл аталғандар жеке мәселелерді шеше алатын ғана бағдарламалар мен мәтіндік файлдар жиынтығы және олар бір-біріне тығыз байланыста болмай,

тұтынушы сұранысына бейімделмегені әрі бір орталықтан басқарылатын комплексті түрдегі автоматтандырылған күрделі жүйе болмауынан деп білеміз.

Жоғарыда сөз болып жатқан Пленум мәжілісінде түркі тілдерінің көпаспектілі машиналық қорын жасау қолданбалы ілімдердің мұқтаждығына сай жүзеге асырылуы тиіс екендігі баса айтылды.

Өкінішке қарай, мұндай жақсы бастама сөз жүзінде ғана қалып, түркі тілдерінің көпаспектілі машиналық қорын өмірге келтіру әлі де болашақтың ісіне айналып отыр.

Ал қазіргі жағдайда, ең алдымен, егеменді еліміздің мемлекеттік тіліне айналған ана тіліміз – қазақ тілінің компьютерлік қорын ғылыми тұрғыда құру – жаңаша ойлап, жаңаша шешудің ең тиімді жолы деуге әбден-ақ болады.

Енді түркі тілдерінің көпаспектілі машиналық қорын құруға қажетті деп есептелетін түрліше салаларына (блоктарына) жалпылама сипаттама бере отыра, сол идеяны қазақ тілінің компьютерлік қорын жасау кезінде де пайдалануға болады-ау деген оймен, оның төменде аталатын тарамдарын оқырман қауымға ұсынып отырмыз.

1. Алдымен құрастыруды қажет ететін сала – лексикалық блок. Ол орфографиялық, орфоэпиялық, түсіндірме, терминологиялық, жиілік және аударма сөздіктерді қамтитын көпшілік қолды лексикографиялық жүйе болып табылады.

2. Грамматикалық блок – лексикалық және грамматикалық морфемалардың тізбелерінің жиынтығынан, сондай-ақ синтаксистік модельдерді (сызбалық) қамтитын тілдік материалдардан құралады.

3. Фонетикалық блок – тілдің фонемалық құрамын анықтайды, әр фонеманың айырым (дифференциал) белгілерінің анықтамалары беріледі.

4. Морфонологиялық блок:

а) тілдік бірліктердің фонемдік құрылысының ерекшеліктері жөніндегі деректер қамтылады. Мәселен, негіздер, форманттар, сөздердің жекелеген кластары (бейнелеуіш сөздер) т.б.;

ә) тілдің сингармонизм заңдылығы, сондай-ақ дауыссыз дыбыстардың ассимиляциясы, қатаңдауы мен ұяңдауы жайындағы мәліметтер ескеріледі;

б) екпіннің орны көрсетілген түбір негіздер мен сөз-формалар жинақталады.

Міне, осы аталған салалардың бәрін жан-жақты қамти отыра, қажетті деген сұранысқа толық та және тұжырымды жауап алу үшін біз соңғы ғылыми-технологиялық жетістіктерді пайдалана отырып, әрі жүйелі, әрі сапалы түрде көрініс табатын ғылыми тұрғыдағы қазақ тілінің компьютерлік қорын іске қосуымыз қажет.

Ұсынылып отырған жоба іске асып жатқан жағдайда мынадай сұрақ туындауы мүмкін: "Қазақ тілінің компьютерлік қоры" жасалып болғаннан кейін ол тек архив тәрізді қызмет ете ме, әлде ол тіл жөніндегі жаңа білімдер қосатын, жаңа мәліметтер тауып беретін ғылыми зертхана тәрізді болмақ па?

Бұл ретте айтарымыз, екі жақты құбылыстың соңғысы алғашқы қисынның барлық қасиеттерін иемденетіндігі себепті, оның берер мүмкіншіліктері де молырақ болмақ. Демек, "Қазақ тілінің компьютерлік қоры" бізге белгілі тілдік заңдылықтар жөніндегі мәліметтерді бойына жинақтаған тілдің синхронды және диахронды процестерін модельдей алатын әрекет үстінде болуы қажет. Сонда ғана компьютерлік тілдік қор түрлі деңгейдегі тіл қабаттарымен жүйелі негізде қызмет ететін болады. Басқаша айтқанда, "Қазақ тілінің компьютерлік қоры" дегеніміз қазақ тілі жайлы барлық мәліметтерді жинақтаушы банк, бірақ ол "ақылды", мамандандырылған білім қоры. Оны өзімізге дәстүрлі кітап ретінде де пайдалануға болады және

зерттеуші өз ұсыныстарын белгілі бір мақсаттарда жүзеге асыра алады деген сенімдеміз.

Осы саладағы белгілі ғалымдардың тұжырымдауынша, компьютерлік тілдік қор дегеніміз ғылым адамының өз зерттеу нысанына жаңаша тұрғыда көз салудың мүмкіндігі болып табылады. Мұндай тілдік қор неғұрлым қомақты болса, солғұрлым тіл құрылысының сырын тереңірек аша алады, сөйтіп зерттелетін нысан жөніндегі түсініктердің шеңбері кеңиді, адамның білім өрісіндегі "ақтаңдақтардың" бедер-бейнесі айқындала түседі.

Біздер де осы айтылғандарға қосыла келе, өз атымыздан "Қазақ тілінің компьютерлік қоры" зерттеуші адамның қалып-қабілетін әлденеше есе арттырады, шығармашылық қуат көздері ашыла түседі, сөйтіп бұл жаңа мүмкіндіктер қазақ тілінің жүйелілік қасиеттерін жетілдіруге және тіл жүйесін мұқият тануға жұмсалар еді деп тұжырымдар едік. Әрине, бұл аталып отырған тілдік қорды өмірге келтіру үшін оның жоба-жолдары мемлекеттік дәрежеде, яғни ҚР Білім және ғылым министрлігі, Тіл комитеті деңгейінде қаралуы қажет. Себебі, мұндай жалпы филологиялық, ұлттық және мәдениет дәрежесіндегі күрделі жүйені іске қосу – қажетті мамандарды (жоғары дәрежелі инженер-бағдарламашыларды, компьютерлік лингвистика саласының мамандарын және т.б.) және тиісті қаражат мүмкіндігін шешуді талап етеді.

Сөз соңын, машиналық қор жайындағы идеяны осыдан ширек ғасыр бұрын ең алғаш рет ұсынған информатика саласының белгілі ғалымы А.П.Ершовтың сол кездердегі айтқан сөзімен қорытындылағымыз келіп отыр.

А.П.Ершов 1978 жылы жарық көрген "Іс қағаздары прозасының феномены: диалогтар жүйесін жасау методологиясына" атты мақаласында: "Егер орыс тілінің машиналық қорын жасау мақсаты ойдағыдай орындалса, модельдеу мен алгоритмдерді құрастыру саласындғы қол жеткен табыстарымыздың бәрі академиялық жаттығулар қалпында қалып қоймай, іс жүзіне асқан болар еді, мұндай өте-мөте қажетті проблеманың шешілуі ғылым, мәдениет және қолданбалы техникалық тәсілдер тұрғысынан мейлінше құнды болмақ. ...Тілдің машиналық қорын жасау ісі өндірістік лингвистикалық жүйелерді іске қосудан озық болуы тиіс, өйткені тек осы жағдайда ғана толып жатқан қажетсіз қайталаулардан құтылып, тілдің жанды өзегін озбырлар мен дүмшелердің қиянатынан аман сақтап қалуға болады" (*Ершов, 1979*).

Гузев В.Г., Пиотровский Р.Г., Щербак А.М. О создании машинного фонда тюркских языков // Советская тюркология. 1988. № 2. С. 98-101.

Ершов А.П. К методологии построения диалоговых систем: феномен деловой прозы. Препр. № 156. ВЦ СО АН СССР. Новосибирск, 1979.

Машинный фонд русского языка: идеи и суждения. М.: Наука, 1986. 240 с.

ҚОРЫТЫНДЫ

Тілдік құбылыстардың біркелкі еместігі, бір қарағанда, компьютерді тіл білімі саласындағы зерттеу жұмыстарында пайдалануға кедергі келтіретін сияқты. Егер біз компьютердің мүмкіндігін толық түрде ескере білсек, тіпті шешімі жоқ-ау деген көптеген мәселелердің де шешімі табылар ма еді. Көбінде, компьютер жадында орындалатын амалдар тілшінің орындайтын формалды іс-әрекетіне ұқсас және олар көрініс деңгейінде тіл мен сөйлеу аспектілерінің көптеген жақтарын қамти алатыны сөзсіз. Сондықтан бұл жердегі ең бастысы адам баласы өзінің ойлау процесін компьютерге бөлшектеп беріп, формалдау жолдарымен ұштастыра білуі қажет сияқты. Сонда ғана зерттеуші тілдік парадигмалық көріністерді, фразалық тармақтанудың заңдылықтарын ашуды және т.б. формалдануға қажетті жұмыстарды іске асырар ма еді.

Қазақ тіл біліміндегі зерттеу кезендеріне, әсіресе, лингвистикалық ақпараттармен жұмыс істеу: оларды жинақтау, іздеу, сақтау, шет тілдегі ақпараттарды қазақ тіліне аудару, ақпараттық библиографиялық іздеу, қазақ тіліндегі мәтіннен реферат алу және т.б. істерге логикалық роботтың (компьютердің) араласуы жақын болашақта мемлекеттік тұрғыда қайта қаралып жатса, бұл мәселе өндірістік дәрежеге дейін көтерілуі де мүмкін.

Ғылым мен техниканың дамуы тіл қызметінің қоғамдағы коммуникативтік қолданыс аясын кеңейте түскені бәрімізге мәлім. Тіл өзінің адамдар арасындағы қарым-қатынастық қызметін атқаруымен бірге адам мен компьютердің қазақ тіліндегі диалогын да жүзеге асыруы қажет деп білеміз. Осы себептен жақын болашақта табиғи тіл және оның түрлері ақпарат таратудағы **"адам – компьютер – адам"** аралығындағы коммуникативтік жүйе негізін құрайды деп сенім білдіруге әбден болады. Әрине, аталған жүйеде – тіл, өзінің адам баласының ойлау қабілетіндегі үйреншікті ассоциативті-эвристикалық мүмкіншілігінен басқаша, өзіне жат қалыпқа түсіп, компьютердің тізбектік-алгоритмдік субстратына ауысады. Тілдің бұл құбылысы қазақ тіл біліміндегі ізденістерге біршама методологиялық қиыншылықтар туғызуы мүмкін. Осы себептен, қазіргі тілдік зерттеулердің, біріншіден, қатаң логикалық теорияға бағындырылуын, екіншіден, объективті және толық болуын барынша талап ету қажет.

Тіл қызметінің жоғарыда сөз болған шарттары тіл білімінің "инженерлік лингвистика" деп аталатын жаңа саласында қарастырылады.

Инженерлік тіл білімі, негізінен, электрондық машина (компьютер) қажеттігі үшін тілдік белгілерді, тіл жүйесі мен оның атқаратын қызметін (функциясын) модельдеу мәселелерімен айналысады.

Тілдік объектілерді жасанды түрде қайта туындату үшін, тек қана жоғары дәрежедегі техникалық база болуы жеткіліксіз. Біздің ұйғаруымызша, қолданбалы есептер үшін, ең бастысы, күшті және объективті теория құрылуы қажет. Сондықтан да әрбір зерттеушінің тәжірибе жүзінде алған нәтижелері сол лингвистикалық теорияға баға беретіндей өлшемге айналуы керек.

Инженерлік тіл білімінің қажеттігін дәлелдейтін проблемалық бағыт-бағдарлардың ең негізгісі – мазмұн бірлігін автоматты түрде айқындап білу және табиғи тіл мен жасанды тілдерді семиотикалық жағынан барынша жақындастыру проблемалары. Сонымен қатар ашық жүйеге жататын табиғи тіл және келісіммен

алынған, жабық семиотикалық жүйедегі қарым-қатынастарды зерттеу – инженерлік тіл білімінің методологиялық және технологиялық мәселелерінің ең бір негізгілері. Бұл мәселенің дұрыс шешім табуы – семиотика, информатика, компьютерлік бағдарлама жасау және математикалық машиналар теориясы, ойлау процестерін модельдеу салаларындағы ғылымдардың қарқындап дамуының бірден-бір шарты.

Инженерлік лингвистиканың тіл біліміндегі басқа да пәндер қатарынан тең дәрежеде орын алуы – көптеген тілшілер мен математиктердің, инженер-программистердің жүргізген тәжірибелерінің, теориялық ізденістерінің нәтижесі деп түсінген жөн. Мәселен, қазіргі қазақ тілінінің "Статистикалық лингвистика" және "Инженерлік лингвистика" салалары – физика-математика ғылымдарының кандидаты, филология ғылымдарының докторы, профессор Қ.Б.Бектаев есімімен және ол ғалымның шәкірттерінің көптеген ғылыми ізденістерімен тығыз байланысты. Сол сияқты, қазақ тілі грамматикасын формалды сипаттауға байланысты соңғы кездегі ғылыми ізденістерімен көзге түсіп жүрген техника ғылымдарының докторы, профессор А.Ә.Шәріпбаевты және т.б. атауға болар еді. Аталған зерттеу нәтижелері қазақ тілін инженерлік лингвистика әдіс-тәсілдерімен қарастыруда, мемлекеттік тілімізді оқытудың (үйретудің) автоматты түрдегі жүйесін құруда және қазақ тілінің формалды грамматикасының ғылыми негізін құру жолының бастамасы болатынына еш күмәніміз жоқ.

Классикалық лингвистика, көбінде, ақпараттық процестің қандай материалдық ортада жүріп жатқанына мән бере бермейді. Ал инженерлік тіл білімінде басқаша, яғни табиғи тілдің "адам – компьютер – адам" жүйесіндегі қызметінің ерекшеліктері стратегиялық және технологиялық ортаға тікелей қатысты. Сондықтан тілдік мәліметтерді компьютер арқылы қайта қарастыру жағдайында, адам баласының миы мен компьютер жадының құрылымды-функционалдық сипаттарын салыстыра отырып, оларға тән қасиеттерді ескерген жөн. Олай дейтініміз, тіл білімі саласындағы электрондық машиналар мен адам баласының мүмкіншіліктері жайлы көптеген тілшілер мен математиктер ғылыми әдебиеттер бетінде өз ойларын білдіріп жүрсе де, тіл білімі саласындағы компьютердің болашағы жайлы ойлары әлде де бірізді емес. Мысалы, біреулері ЭЕМ-ның жады мен жылдамдығын місе тұтпаса, ал басқалары – компьютер арқылы семантика-синтаксистік, тіпті стильдік ерекшеліктерге қатысты күрделі-күрделі мәселелердің болатынын сөз етіп, қалай болғанда да компьютерді тілден алшақтатуға тырысатындар да жоқ емес.

Қазір, компьютерлік техниканың барынша дамыған кезінде, оның жадының көлемі, жылдамдығы туралы сөз болмаса да, басқа ғылыми мәселелердің "кезек күттірмейтін" тұстары (мысалы, түрлі сөздіктер түзу) себеп болып, қазақ тілін зерттеуде компьютерлік тіл білімінің жолы ойлағанымыздай болмай, қарқыны бұрынғыдан да бәсеңсігенге ұқсайды.

Адам мүмкіншілігіне қарағанда, компьютер, сырт кедергіге қарамай, жадындағы ақпаратты әрі тұрақты, әрі берік, әрі ұзақ мерзімге сақтайтыны және қажетті ақпаратты тез арада қағаз не монитор бетіне шығарып бере алатындығы көпшілікке мәлім. Ал мәтін ішіндегі ақпаратпен (мысалы, мағына мен мазмұн) жұмыс жасауда адамның ми қызметі оның физиологиялық және психикалық қалпына байланысты болатын болса, мұндай қасиет компьютерге жат деуге болады.

Инженерлік тіл білімінің даму болашағын сөз еткенде, мысалы, жақын он жылдың ішіндегі зерттеудің стратегиясын анықтауда және үйлесімді технологияны таңдауда, тек қана бүгінгі күнгі кибернетикалық техниканың және белгілі шеңбердегі тіл білімі мен нейрофизиологияның ғылыми жетістіктеріне сүйенген жеткіліксіз деп есептейміз. Бұл жерде кибернетикалық техниканың даму қарқынын ескере

отырып, кибернетиканың, тіл білімінің және нейрофизиологияның методикалық жақтарын ескеретін кең шеңбердегі алдын болжай алатындай ғылыми көзқарас қажет.

Қазіргі кезде еуропа тілдеріндегі мәтіндік хабарламалардың грамматикалық, сөздік, фразеологиялық, кейбір лексика-грамматикалық және, тіпті, семантикалық деңгейлерін айқындайтын компьютерге арналып жасалған алгоритмдер негізіндегі бағдарламалары іске қосылған.

Осы айтылғандардың негізінде төмендегідей сұрақ туындауы мүмкін.

Электронды автоматтың (компьютердің) болашақтағы мәтінді түсіндіру мүмкіндігі қай деңгеймен шектелмек? Ол тек лексика-грамматикалық, лингвистикалық деңгей ме? Әлде компьютердің мәтін мазмұнын анықтаудағы мүмкіндігі маман аудармашының деңгейіне теңелуі мүмкін бе? Компьютерлік аударма проблемасына қатысты бұл сұрақты былай да қоюға болады: шетел тіліндегі белгілі бір сала бойынша алынған мәтінді әрі осы саланың маманы, әрі тәжірибелі аудармашы болып табылатын адамның аударуы мен компьютерлік аударманы салыстыруға бола ма? Әлде компьютерлік аударма еш уақытта жоғары сапалы болмай, оған тұтынушының амалсыз араласуы қажет пе?

Әрине, қазіргі кездегі компьютерлік аудармалар әлде де маман аудармашының жәрдемін қажет етеді. Сондықтан да кибернетикалық автоматтың мәтіндерді "адамша" аударуы бір адамның бойына біткен лингвистикалық білім дәрежесімен және басқа да табиғи мүмкіншіліктерімен шектелмеу керек, ол компьютер үшін жеткіліксіз. Болашақтың лингвистикалық компьютері бүкіл бір ұйымға тән, әлеуметтік кәсіби (профессионал) топтарға тән интеллектуалдық, мәдени-әлеуметтік және лингвистикалық тәжірибелерді өзіне игеруі қажет.

Сонымен, хабардың берілу жүйесінің салалық тілдермен жұмыс істеуі, табиғи тіліміздің коммуникативтік қызметін дамыта түсетіндігіне еш күмән жоқ деуге болады. Ал қазақ тіл білімінің компьютерлік саласының тиісті дәрежеде дамуын, оның қатаң ережеге бағынуынан, мәтін бойындағы тілдік бірліктерді тануға қажетті формалды қасиеттерінің айқындалуынан, омоним сөздер мен көпмағыналы сөздердің және синоним сөздер қатарын сирету мен реттеу мәселелерінің өз шешімін дұрыс табуынан байқауға болар еді.

МАЗМҰНЫ – СОДЕРЖАНИЕ

СЫЗДЫҚ Р., ӘНЕС Ғ. Қазақ тілтанымының жаңа бір саласының көшбастары 7
СЫЗДЫК Р., АНЕС Г. Основатель нового направления казахского языкознания 7

Қолданбалы лигвистика: Қазақ тілінің статистикасы 15

Основные принципы формализации содержания казахского текста 157

Қолданбалы лингвистика: Формалды модельдер 347

Компьютерлік лингвистикаға кіріспе 515

"Арыс" баспасының директоры
Ғарифолла ӘНЕС

Асқар ЖҰБАНОВ

ҚОЛДАНБАЛЫ ТІЛ БІЛІМІНІҢ МӘСЕЛЕЛЕРІ

ВОПРОСЫ ПРИКЛАДНОГО ЯЗЫКОЗНАНИЯ

Редакторы
Ғарифолла ҚАБДОЛҚАЙЫРҰЛЫ

Тех. редакторы
Саят ТӨРЕБАЙҰЛЫ

Суретші
Орынбек ӘНЕСОВ

ИБ № 666
Терімге 01.10.2007 жіберілді.Басуға 31.11.2007
қол қойылды. Пішімі 70x100 $^{1}/_{16}$. Офсеттік қағаз.
Шартты баспа табағы 60,0. Есепті баспа табағы 40,0.
Таралымы 2000. Тапсырыс № 1184.

050009, Алматы қаласы,
Абай даңғылы, 143. 5-қабат.
"Арыс" баспасы
тел./факс: 242 65 56
E-mail: arys@nursat.kz
www.arys.kz

Тапсырыс берушінің файлдарынан Қазақстан Республикасы
«Атамұра корпорациясы» ЖШС-нің Полиграфкомбинатында басылды.
050002, Алматы қ., М. Мақатаев к., 41.

Ардақты Алаш азаматтары!

Репрессияға ұшыраған Қазақстан зиялыларының мұрасын зерттейтін «АРЫС» қоры он бес жылдан бері жазықсыз жазаланып, аттары аталмай, көпшілік жадынан өше бастаған біртуар азаматтарымыздың мұрасын жинап, еңбектерін жарыққа шығарады; ескерткіш, мүсіндерін жасап, музейлерін жабдықтайды; арыстарымыздың өмірі мен творчествосын насихаттаушыларға жәрдем беріп, мерейтойларын өткізуге атсалысады; ардақтыларынан ажырап қалған шаңырақтарға қайырымдылық көмек көрсетеді.

Қор жанындағы «АРЫС» баспасы Сіздің түрлі мәдени тапсырыстарыңызды тез әрі сапалы орындап береді

050009, Алматы қ.
Абай даңғылы, 143
5-қабат.
527, 529-бөлмелер
тел./факс: 242 65 56

АРЫС

е/ш.: 001700102
30070492 (валюта)
Алматы қ-сы
Казкоммерцбанк
МФО 190501724
РНН 600400010767

www.arys.kz
E-mail: arys@nursat.kz

Дорогие читатели!

Фонд "АРЫС", созданный ведущими казахстанскими филологами, историками, политологами и публицистами, более пятнадцати лет занимается сбором, исследованием и публикацией наследия передовой казахской интеллигенции, ставшей жертвой политических репрессий 20 века. Фонд стремится восстановить объективную историческую память: выпускает книги, содействует открытию музеев и оформлению тематических экспозиций, участвует в проведении юбилеев выдающихся деятелей, оказывает посильную помощь их потомкам.

Издательство при фонде "АРЫС" квалифицированно и качественно выполнит заказы, относящиеся к сфере науки и культуры, выпустит Ваши литературные труды, брошюры, авторефераты, буклеты и др.