

12013  
42326к

АСКАР  
КУДАЙБЕРГЕНУЛЫ  
ЖУБАНОВ

**Основные принципы  
формализации  
содержания  
казахского  
текста**



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РЕСПУБЛИКИ КАЗАХСТАН  
НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК  
ИНСТИТУТ ЯЗЫКОЗНАНИЯ ИМ. А.БАЙТУРСЫНОВА**

**Аскар Кудайбергеноулы  
Жубанов**

**Основные принципы формализации  
содержания казахского текста**

**Алматы  
2002**

ББК 81.2

Ж 83 ✓

**Жубанов Аскар Кудайбергенолы**

**Ж 83 Основные принципы формализации содержания  
казахского текста. – Алматы, 2002. – 250 с.**

**ISBN 9965-13-426-X**

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.

Разработанный автором способ статического и динамического представления содержания казахского текста может быть применен при составлении учебников и учебных пособий.

**ББК 81.2**

**Ответственный редактор**

**доктор филологических наук З. М. Базарбаева**

**Рецензенты:**

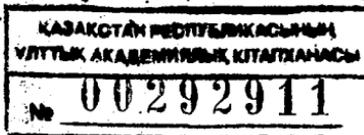
**доктор филологических наук профессор З. К. Ахметжанова;**

**доктор филологических наук профессор С. М. Мырзабеков.**

Монография рекомендована к изданию Ученым советом Института языкознания им. А. Байтурсынова Министерства образования и науки Республики Казахстан.

Ж  $\frac{4603000000}{00(05) - 02}$

СМ



**ISBN 9965-13-426-X**

© . Жубанов А. К. 2002

*Преклоняясь перед гениальностью и мужественностью Отца Кудайбергена Куановича Жубанова – филолога-мыслителя, классика казахского языкознания, ставшего жертвой репрессий 1937 года, посвящаю . . .*

## 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, неоднозначность, 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 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 нашей 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 индивидууму 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 Д. 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 впечатков, 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 пони-

мать “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, предло-

женного 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 искусственного 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 также 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

яности 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, стохастичность 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 последова-

тельность 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 внесязыковую 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 обуславливает понятие слов определенной тематической группы на протяжении всего текста.

Помимо разнообразных дейктических элементов и различных средств собственно синтаксической связи (союзы, ссоюзные 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 presupпозиций [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 зна

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 преопределили 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 осо-

бого 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 антimentsализм. 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, средствами отражения объективной реальности и т.д.

Следующее свойство знака связано с тем, что форма знака в знаковых системах конвенциональна, 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. Конвенциональность 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 рядом 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 (интенционал, 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 с соавт. вводят в структуру знака третьему компоненту – коннотат, который над денотатом и десигнатом каждого знака надстраивает вторичные смысловые оттенки и оценки – эмоционально-оценочные аспекты