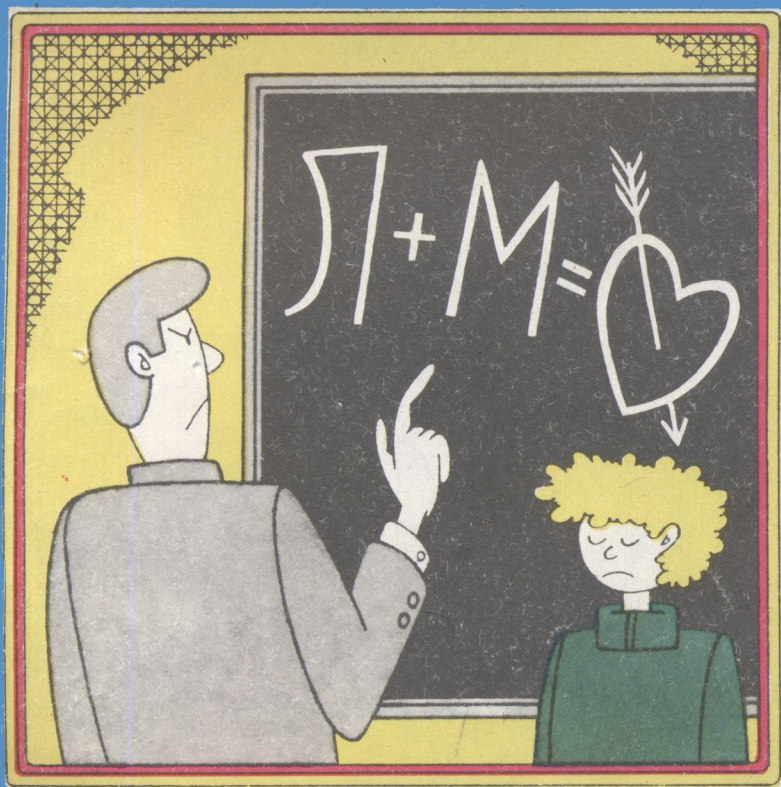


# МР И знаний

Г.Е. КРЕЙДЛИН А.Д. ШМЕЛЕВ

## Математика помогает лингвистике







Г. Е. КРЕЙДЛИН А. Д. ШМЕЛЕВ

# Математика помогает лингвистике

*Книга для учащихся*

Серия «Мир знаний» основана в 1972 г.

Данное издание представляет собой авторскую работу, вошедшую в число победителей в открытом конкурсе, который реализуется совместными усилиями

Министерства образования России,  
Государственного комитета РФ по высшему образованию,  
Международного фонда «Культурная инициатива» и  
Международной ассоциации развития и  
интеграции образовательных систем.

Конкурс является составной частью программы  
«Обновление гуманитарного образования в России».

**Спонсором программы является  
известный американский предприниматель и  
общественный деятель Джордж Сорос.**

#### Рецензенты:

доктор филологических наук, ведущий научный сотрудник  
Института языкознания АН СССР *Л.П.Крысин*;  
доктор физико-математических наук, профессор *А.В.Гладкий*;  
учитель русского языка школы № 57 Москвы *Е.В.Вишневецкая*

Художник *А.Е.Шабельник*

**Крейдли Г.Е., Шмелев А.Д.**

**К 79** Математика помогает лингвистике: Кн. для учащихся. — М.: Просвещение, 1994. — 176 с.: ил. — (Мир знаний). — ISBN 5-09-004370-1.

В книге в доступной форме раскрывается сходство и различие русского языка и языка математики, описываются математические методы, используемые при изучении фонетики, орфографии, лексики, словообразования, морфологии, синтаксиса, стилистики. Предлагаются лингвистические задачи, вопросы для размышления, ответы на них.

Книга предназначена для школ с углубленным изучением математики, а также для учащихся гуманитарных классов, лицеев и гимназий. Книга может быть использована на факультативных занятиях.

К 4306020000 — 519

103(03) — 94

16 — 93, III—IV кварт. 1993 г.

ББК 81

ISBN 5-09-004370-



Scan AAW

© Крейдли Г.Е., Шмелев А.Д., 1994

## ОТ АВТОРОВ

Язык, математика... Сколько раз мы сталкивались с этими словами, противопоставляя их друг другу! «Русский язык — это интересно, а математика — ничего не поймешь!» или «Математика — это да, а русский я терпеть не могу».

И вот мы открываем книгу, где слова эти, язык и математика, не только не противопоставлены, а как бы соединены друг с другом: ведь в заглавии говорится о помощи одной науки другой. Но разве можно соединить несоединимое? Как может помочь такой тонкой и неосязаемой материи, как язык, строгая, сухая, наполненная формулами и значками математика?

В этой книге вы узнаете о возможных применениях математических методов в ярко выраженной гуманитарной области — языкознании. Мы познакомим вас с историей идей, с мнениями и раздумьями ученых над проблемами формализации языка, покажем результаты их исследований в области создания так называемых интеллектуальных систем, таких, как «думающие машины», автоматические системы обработки информации, системы машинного перевода, системы шифровки и дешифровки текстов и многие другие.

Попутно расскажем о том, что применительно к науке о языке означают много раз встречавшиеся вам слова *формальный* и *неформальный*, что такое строгое рассуждение и чем оно отличается от нестрогого, что такое формальное определение в гуманитарных науках. Наконец, мы изложим идею нескольких алгоритмов нового и, вероятно, неизвестного вам типа — лингвистических алгоритмов — и рассмотрим разнообразные виды задач, которые они призваны решать.

Говоря о применении математических методов при изучении языка, мы, разумеется, не имеем в виду обучить вас говорить и писать в повседневной жизни на языке математики. Речь пойдет об использовании математики в науке о языке — языкознании, или лингвистике, — о точных методах решения возникающих в лингвистике проблем.

Сразу предупредим вас: чтение этой книги будет делом совсем не простым, ведь речь пойдет о взаимодействии двух сложных дисциплин. Мы постараемся облегчить вам чтение, разбирая много примеров, давая неформальные пояснения,

помещая рисунки и таблицы, привлекая разнообразные аналогии.

В конце книги расположены ответы или пояснения к вопросам и задачам, которые предлагаются читателю по ходу изложения материала. И задачи и ответы обозначены особым рисунком:



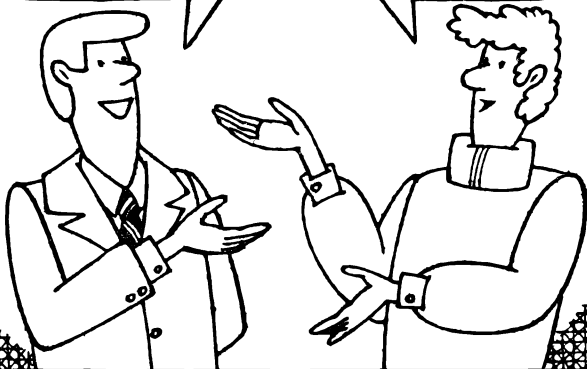
И, конечно, мы будем рады, если вас заинтересуют точные методы исследования в гуманитарных областях, если вы вообще научитесь о самых разных объектах рассуждать строго и логично, если сможете проникнуться атмосферой научных поисков и научных дискуссий, сопровождающих создание всякой новой научной дисциплины.

Глава первая

# Существует ли язык математики?

- Множество  
М бесконечно.

- Я приобрёл  
множество  
друзей!





# І. ЯЗЫК И ЯЗЫКИ

## 1. ПОДЪЯЗЫКИ

**Я**зык и математика... Вы можете спросить: о каком же языке идет речь — о русском, украинском, английском, немецком? А ведь еще есть бенгальский, греческий, корейский... К какому же из них применяется математика?

Вообще говоря, математические методы как совокупность приемов описания универсальны в том смысле, что их можно применить к любому из естественных, «человеческих», языков, хотя в дальнейшем мы будем опираться на факты русского языка.

Между тем русский язык (как, впрочем, и любой другой) представляет собою весьма сложное и неоднородное образование. Когда мы читаем в книге по биологии: «В клетке можно выделить три области: клеточное тело, дендриты и аксоны... На клетке может быть также до 10 000 синапсов» — или в статье об экономике: «Прямое банковское кредитование строительных организаций», — мы либо вообще ничего не понимаем, либо понимаем очень мало. Во всяком случае ясно, что перед нами не совсем тот (или даже совсем не тот) язык, которым мы пользуемся в повседневном общении друг с другом. Фактически в пределах русского языка мы имеем дело как бы с разными языками: языком биологии, языком экономики, юридическим языком. Говорят также о языке улицы, языке радио, языке деловых бумаг и многих других языках. Открыв томик стихов известного поэта А.Яшина, читаем в одном из стихотворений: «Деука на лауке верёуку плетет». Фраза демонстрирует нам, как говорят на родном языке автора. Это еще один язык — язык его родной Вологодской области. В книге Л.Успенского «Слово о словах» мы встречаемся с такой фразой: «Батька уже помешался, так ен на будворище орет, а матка — тая шум с избы паше». Приведенная фраза, как ни странно, тоже произнесена на русском языке (на псковском диалекте), но вам, чтобы понять ее, наверно, нужно увидеть ее перевод... на русский же язык. Пожалуйста: «Отец уже вспахал поле и теперь пашет на огороде, а мать подметает избу».

Итак, как мы видим, разновидностей русского языка очень много. Такие разновидности принято называть **п о д ъ -**  
**я з ы к а м и**. На практике всякий человек владеет сразу несколькими подъязыками, используя их в разных ситуациях. Одним он пользуется в разговоре с друзьями, другим — когда



пишет официальную бумагу (например, объяснительную записку или жалобу). Таким образом, каждый человек в каком-то смысле полиглот, даже если он не знает иностранных языков.

А как человек овладевает всеми этими подъязыками? Одни даны ему как бы с рождения, а другим он специально обучается (например, в школе). Придя на урок биологии, мы знакомимся с языком биологии; на уроках физики — с языком физики. Стать специалистом в какой-либо науке и значит, среди прочего, в совершенстве овладеть ее языком. И не будет преувеличением сказать, что на всех уроках в школе, а не только на уроках русского или иностранного языка, мы изучаем языки.

## **2. ЯЗЫК МАТЕМАТИКИ КАК ПОДЪЯЗЫК ЕСТЕСТВЕННОГО ЯЗЫКА**

Теперь мы остановимся подробнее на некоторых важных особенностях одного подъязыка — языка математики. При этом мы будем сопоставлять его с «обычным» русским языком повседневного общения. Сразу же оговоримся, что строгой границы между этими двумя языками нет: не существует критерия для определения принадлежности любой конкретной фразы тому или другому подъязыку.

Однако есть отдельные признаки, присущие только языку математики, но не «обычному» языку. О математическом языке говорят как о «сухом», лишенном выразительности, неэмоциональном. Это ощущение «сухости» возникает из-за



сравнительно небольшого запаса используемых синтаксических конструкций, из-за отсутствия эмоционально окрашенной лексики, монологического характера речи. Лексические и синтаксические средства математического языка отличаются однозначностью, отсутствием подтекста, намеков, строгой определенностью основных понятий. В математических текстах встречаются особые знаки — символы, индексы, формулы, которые редко встретишь в языке повседневного общения.

Приведем характерный пример математического текста — текста задачи: «На прямой дано более чем одноэлементное множество точек  $M$ , такое, что каждая точка из  $M$  является серединой отрезка, соединяющего две другие точки из  $M$ . Докажите, что множество  $M$  бесконечно».

Начнем с того, что многие слова этого текста вообще не встречаются в бытовом языке или же встречаются в других значениях: *множество, точка, прямая, середина, отрезок, бесконечно*. Здесь эти слова начисто лишены эмоциональной окраски, тогда как в разговорной речи те же самые слова очень часто бывают эмоционально насыщены и полны выразительности: «Все кончено. *Точка!*», «Вы мне *бесконечно* дороги», «Где ты приобрел такое *множество* друзей?». Нас совсем не удивляет отсутствие междометий, звукоподражаний, существительных и прилагательных с уменьшительно-ласкательными суффиксами. Синтаксис текста задачи также хорошо отражает особенности математического языка. Это использование заглавной буквы латинского алфавита в качестве приложения (а в дальнейшем — в качестве самостоятельного

имени), наличие особых конструкций, свойственных языку математики (*на прямой дано множество; такое, что; докажите, что...*), отсутствие восклицательных предложений, прямой речи, диалогов.

### 3. ПОНЯТИЕ МЕТАЯЗЫКА

Мы не случайно заговорили о различных подъязыках. Дело в том, что к одним из них формальные методы исследования применять уместно, и сделать это не очень сложно, а другие с трудом поддаются какой-либо формализации. Например, язык деловых документов легко описывается формальными средствами (и понятно почему), а язык художественной прозы допускает формальное описание лишь в очень немногих своих аспектах. Представляется, что математический подъязык может быть описан математическими методами, во многом благодаря тем его особенностям, о которых мы только что говорили.

Получается, что один и тот же подъязык используется и как средство, с помощью которого мы строим формальное описание, и как объект этого же описания (т.е. именно то, что описывается). Так часто бывает в науке о языке: изучая какой-то язык (или его фрагмент), лингвисты сознательно прибегают к этому же языку как удобному инструменту описания. Вспомним уроки русского языка. Всякое правило, фиксирующее некоторое языковое явление, формулируется на русском же языке. Мы этого, как правило, не осознаем, но тем не менее при строгом, формальном описании нам приходится последовательно различать как бы две стороны русского языка: языка-объекта и языка-инструмента (или, как говорят, м е т а я з ы к а). Неразличение двух указанных сторон может привести к ошибкам. Ученик седьмого класса, разбирая предложение *Как бы подошел здесь этот глагол!*, назвал слово *глагол* глаголом, очевидно, спутав существительное языка-объекта *глагол* с термином языка-инструмента. Такого рода ошибок можно избежать, если использовать в качестве метаязыка формальный математический язык.

В дальнейшем язык математики используется в книге в функции инструмента описания, а объектом описания являются различные подъязыки русского языка. Мы не всегда будем указывать эти подъязыки, полагая, что во всех необходимых случаях читатель сам сможет определить, о каком конкретном подъязыке идет речь.

Наряду с математическим языком в функции метаязыка в книге применяется и так называемый а л г о р и т м и ч е с к и й п о д ъ я з ы к. Об устройстве этих двух метаязыков мы расскажем ниже.

## II. МАТЕМАТИЧЕСКИЙ И АЛГОРИТМИЧЕСКИЙ МЕТАЯЗЫКИ

### 1. СЛОВАРЬ

Прежде чем использовать некоторый язык в роли инструмента, его следует описать в явном виде. Тем самым складывается парадоксальная ситуация: до того как стать инструментом, язык должен некоторое время побыть объектом описания. Так в жизни бывает всегда. Например, прежде чем молотку стать инструментом, он должен быть изготовлен, т.е. послужить объектом работы мастера.

Начнем с того, что укажем основные части, из которых состоит всякий язык. Это **с л о в а р ь** и **г р а м м а т и к а**. Словарь и грамматику можно образно представить себе как книги, в совокупности своей содержащие основные сведения о языке.

В словаре представлены слова языка и при каждом слове приводятся сведения о его значении и особенностях употребления. В грамматике даются правила сочетания слов, а также правила образования новых слов, словосочетаний и предложений.

Теперь мы перейдем к описанию наших двух метаязыков: математического и алгоритмического.

Словари этих языков имеют много общего, настолько много, что можно считать оба языка имеющими один и тот же словарь. Что же представляет собой этот словарь? Как и во всяком словаре математического (и вообще формального)



языка, в нем нет междометий, эмоционально окрашенных слов, образных слов и выражений, но зато в нем заметное место занимают специальные термины. Остановимся на них более подробно.

Все термины из словаря наших метаязыков, используемые в данной книге, можно условно разбить на два класса. Во-первых, это слова, относящиеся к общелогическим и математическим понятиям: *множество, элемент, функция, число, операция, последовательность, алгоритм* и др. Во-вторых, это термины, обозначающие универсальные понятия науки о языке: *язык, подязык, знак, значение, словарь, грамматика, говорящий, высказывание* и др.

Описывая все эти слова, мы будем особо останавливаться на тех моментах, которые существенны при обсуждении математических методов, применяемых в языкознании. Изложение мы построим по следующему плану. Дадим (или напомним) определения для определяемых терминов и неформальные (содержательные) пояснения для неопределяемых. После каждого определения или пояснения приводятся примеры.

### 1.А. Множество

**Понятие о множестве.** Начнем с понятия множества. Это неопределяемое понятие, такое же, как натуральное число, точка, плоскость, алгоритм и др. Его смысл можно пояснить так: множество — это совокупность, или собрание, предметов, рассматриваемых как единый ансамбль. Конечно, написанный только что текст нельзя считать определением, поскольку слова *совокупность, собрание* и *ансамбль* означают не более простые понятия, чем «множество». Поэтому нам остается показать на примерах, как употребляется термин *множество*. Так, можно говорить о множестве людей в данной комнате, множестве книг в книжном шкафу, множестве всех целых чисел. Говорят, что множество состоит из элементов, а про эти элементы — что они входят в данное множество, или принадлежат ему. Если элемент  $x$  входит в множество  $A$ , то символически это обозначается:  $x \in A$ . Выражение  $x \notin A$ , как можно догадаться, обозначает, что  $x$  не принадлежит  $A$ .

Математикой предъявлено к множествам два требования: 1) требование различимости элементов и 2) требование несущественности порядка следования элементов. Конечно, эти требования нужно пояснить.

Требование различимости элементов означает, что все элементы множества различны. Давайте посмотрим, что произойдет, если мы не примем во внимание это требование. Представим себе, что на уроке русского языка учитель задает



вопрос: «Сколько букв в слове *касса*?» (Такой вопрос может быть задан, например, тогда, когда ученик ошибся в правописании данного слова и учитель хочет указать ему на ошибку.) Очевидно, правильный ответ на этот вопрос — «Пять». Теперь представим себе, что учитель русского языка заболел и на замену пришел математик. Он мог бы задать, казалось, тот же самый вопрос в привычной для него математической форме: «Сколько элементов содержит множество букв слова *касса*?» Давайте подумаем, правильно ли ответить на этот вопрос: «Пять». С математической точки зрения, т.е. с учетом требования различимости элементов, правильный ответ должен быть другой: «Три, а именно: *к*, *а*, *с*». Ответы на эти вопросы были бы одинаковы только в том случае, если бы ни одна буква в слове, о котором спрашивается, не повторялась.

Требование несущественности порядка следования элементов означает вот что: множество не изменится, в каком бы порядке ни записывать его элементы. Например, слова *вол* и *лов* различны, но множество букв слова *вол* совпадает с множеством букв слова *лов* и может быть представлено как состоящее из трех элементов: *о*, *л*, *в*. Аналогично, множество букв слова *кабан* такое же, как и у слова *банка* или *банк*, а именно: *а*, *б*, *к*, *н*.

Чтобы изучать множества, надо уметь их, во-первых, как-то обозначать, а во-вторых, как говорят математики, задавать.

Множества задаются одним из двух способов: перечислением элементов или указанием свойства, которым обладают

все элементы данного множества, и только они. На вопрос: «Саша, кто у тебя будет на дне рождения?» — Саша может ответить так: «Петя, Вася, Маша, Женя и Лена». Здесь множество гостей Саши задано перечислением имен. Возможен и другой ответ на этот вопрос: «Все мои товарищи из класса». Здесь то же самое множество обозначено по-другому, а именно указанием свойства («быть товарищем Саши из класса»), объединяющего всех перечисленных ребят.

Для обозначения множества в математике обычно используют фигурные скобки:  $\{ \}$ . Обозначение  $\{ \text{Петя; Вася; Маша; Женя; Лена} \}$  читается: «Множество, состоящее из (пяти) элементов: Петя, Вася, Маша, Женя и Лена». Если множество задано через указание свойства, то его элементы вводятся с помощью какой-либо переменной, рядом с которой пишется имя этого свойства, например:  $\{x|x \text{ — товарищи Саши по классу} \}$ . Такая запись читается: «множество таких  $x$ , что  $x$  — товарищ Саши по классу». Перечислением, как правило, задаются конечные множества, т. е. множества, состоящие из конечного числа элементов, а указанием свойства можно задавать как конечные, так и бесконечные множества. Например, множество всех четных чисел задается следующим образом:  $\{x|x \text{ — целое число, делящееся на } 2 \}$ .

Если множество содержит  $n$  элементов, то оно называется  $n$ -элементным. Так, множество  $\{ \text{Петя; Вася; Маша; Женя; Лена} \}$  является пятиэлементным, множество решений уравнения  $x - 2 = 0$  является одноэлементным и может быть записано как  $\{2\}$ . Очень важно не смешивать одноэлементное множество с его элементом. Например, театр, состоящий из одного актера, — это не то же самое, что этот актер.

В математике, как вам должно быть известно, рассматривается также множество, не содержащее ни одного элемента. Такое множество называется пустым и обозначается знаком  $\emptyset$ . Пустое множество может быть задано разными языковыми выражениями:

— множество слов русского языка, начинающихся с мягкого знака;

— множество четырехугольников с пятью вершинами;

—  $\{x|x \in R \text{ и } x^2 < 0\}$ .

Нетрудно понять, что во всех этих случаях, несмотря на различия в языковом оформлении, мы задаем одно и то же множество, не содержащее ни одного элемента, т. е. пустое.

В математике часто говорят также о так называемом универсальном множестве элементов данного типа. Это как бы самое большое множество — множество всех элементов данного типа. Основное его свойство заключается в



том, что какой бы объект данного типа ни взять, он принадлежит этому множеству. Обычно универсальное множество обозначается знаком  $\cup$ .

А теперь давайте проверим себя, все ли здесь мы поняли. Попробуйте-ка ответить на некоторые каверзные вопросы и решить некоторые задачи.



1. Сколько элементов в множестве  $\{\{a\}; a\}$ ?
2. Сколько элементов в множестве  $\{\emptyset\}$ ?
3. Придумать слово русского языка, в котором число букв было бы на три больше, чем число элементов во множестве букв этого слова.
- 4\*. Пусть  $x$  — произвольный русский текст, не содержащий иностранных слов и выражений. Обозначим через  $alph(x)$  функцию: число элементов во множестве букв текста  $x$ . Найти наибольшее значение этой функции.

**Отношения между множествами.** До сих пор речь шла о множествах, каждое из которых рассматривалось изолированно, вне его связи с другими множествами. Сейчас мы, наоборот, остановимся на некоторых важных связях, которые могут существовать между двумя множествами.

Множество  $A$  называется **подмножеством** (частью) множества  $B$ , если любой элемент множества  $A$  входит также и в множество  $B$  (принятое обозначение  $A \subseteq B$ ; здесь знак  $\subseteq$ , построенный по аналогии со знаком  $\leq$  для чисел, называется знаком «нестромого включения»).

С отношением «быть подмножеством» вы все время сталкиваетесь на уроках математики и информатики: вы, несомненно, знаете, что множество натуральных чисел является подмножеством множества целых чисел, а множество букв русского алфавита — подмножеством множества литер. Когда нам объясняют, что волки — это хищники, то мы должны понимать, что множество волков является подмножеством множества хищников.

Принято также считать, что пустое множество является подмножеством любого множества, в том числе и самого себя. Заметим, кстати, что для любого непустого множества  $A$  верно, что  $A \subseteq A$ .

Если  $A \subseteq B$  и  $B \subseteq A$ , то говорят, что множество  $A$  и множество  $B$  равны (или совпадают). Очевидно, что в этом случае  $A$  и  $B$  состоят из одних и тех же элементов, т.е. фактически представляют собою одно и то же множество. Этот факт обозначается как  $A = B$ .

Приведем несколько примеров совпадений множеств: 1)  $\{1; 2\} = \{2; 1\}$ ; 2)  $\{0\} = \{x | x^2 \leq 0\}$ ; 3)  $\{x | x \text{ — слово русского языка, начинающееся с мягкого знака}\} = \{x | x \text{ — четырехугольник с пятью вершинами}\} = \emptyset$ .

Если  $A \subseteq B$  и  $A \neq B$  (читается « $A$  не равно  $B$ »), то говорят, что  $A$  — это собственное подмножество  $B$  (обозначение:  $A \subset B$ ; знак  $\subset$  называется знаком строгого включения). Приведем пример известных вам двух множеств  $A$  и  $B$ , таких, что  $A \subset B$ :  $A = \{x | x \text{ — равносторонний треугольник}\}$ ,  $B = \{x | x \text{ — треугольник}\}$ .

Предлагаем решить несколько задач.



5. Дано:  $A$  — множество простых чисел — делителей числа 30.

$B$  — множество простых чисел — делителей числа 60. Верно ли, что а)  $A \subseteq B$ ; б)  $B \subseteq A$ ?

6. Доказать, что если  $A \subseteq B$  и  $B \subseteq C$ , то  $A \subseteq C$  ( $A, B$  и  $C$  — множества).

7. Доказать или опровергнуть следующее утверждение: «Множество согласных звуков слова *город* является подмножеством звонких согласных звуков русского языка».

**Действия над множествами.** Наверно, у читателя уже возникли вопросы: а зачем нам нужны множества и что мы можем с ними делать? Сейчас мы частично ответим на них, а именно покажем, как по двум данным множествам можно построить третье множество, являющееся результатом соответствующего действия. В этом смысле действия (операции) над множествами очень похожи на арифметические: действие, например, сложения состоит в том, что по двум данным числам ищется третье, которое является их суммой. Мы рассмотрим три основных действия над множествами: пересечение, объединение и вычитание, каждый раз определяя результат соответствующих действий.

1. **Объединение** двух множеств называется множеством, состоящее из всех элементов, которые входят хотя бы в одно из исходных множеств. Если какой-либо элемент  $a$  входит в оба множества, то в объединение он включается один раз.

Обозначение:  $A \cup B$ .

Приведем несколько примеров.

$$1) \{1; 2; 4\} \cup \{3; 5\} = \{1; 2; 3; 4; 5\}.$$

$$2) \{1; 2; 3\} \cup \{2; 3; 4\} = \{1; 2; 3; 4\}.$$

$$3) \{x | x \text{ — звонкий согласный звук}\} \cup \{x | x \text{ — глухой согласный звук}\} = \{x | x \text{ — согласный звук}\}.$$

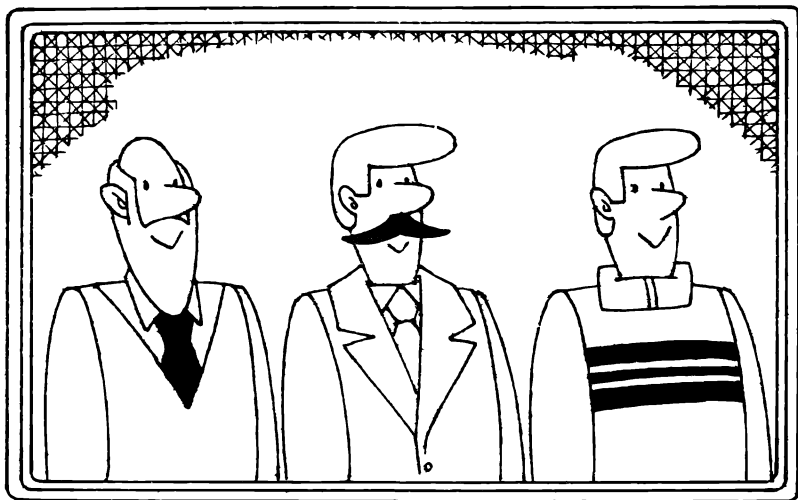
$$4) [2; 3] \cup [3; 4] = [2; 4]^*.$$

$$5) \{x | x \text{ — мальчик}\} \cup \{x | x \text{ — девочка}\} = \{x | x \text{ — ребенок}\}, \text{ или в сокращенной записи } \text{МАЛЬЧИКИ} \cup \text{ДЕВОЧКИ} = \text{ДЕТИ}.$$

6) Рассмотрим текст известной загадки: «Два отца, два сына, а всего трое. Как такое может быть?» Смысл этой

---

\*С обозначением  $[a, b]$ , где  $a$  и  $b$  — числа, вы знакомились на уроках алгебры:  $[a, b] = \{x | a \leq x \leq b\}$ . Такое множество называется числовым отрезком.



загадки в том, что в ней осуществляется не арифметическое сложение, а объединение двух двухэлементных множеств, имеющих общий элемент. Ведь человек, являющийся отцом, может одновременно быть и сыном. Таким образом, в результате объединения получаем трехэлементное множество.

2. Пересечением двух множеств называется множество, состоящее из тех элементов, которые входят в оба исходных множества. Обозначение:  $A \cap B$ .

Примеры:

1)  $\{1; 2; 3\} \cap \{4; 5\} = \emptyset$ .

2)  $\{1; 2; 3\} \cap \{2; 3; 4\} = \{2; 3\}$ .

3)  $[2; 3] \cap [3; 4] = \{3\}$ .

4)  $\text{ХИЩНИКИ} \cap \text{ПТИЦЫ} = \text{ХИЩНЫЕ ПТИЦЫ}$ .

3. Разностью двух множеств  $A$  и  $B$  называется множество, состоящее из всех элементов множества  $A$ , не входящих в множество  $B$ . Обозначение:  $A \setminus B$ .

Примеры:

1)  $\{1; 2; 3\} \setminus \{4; 5\} = \{1; 2; 3\}$ .

2)  $\{1; 2; 3\} \setminus \{2; 3\} = \{1\}$ .

3)  $\{1; 2; 3\} \setminus \{1; 3\} = \{2\}$ .

4)  $\{x | \sin x = 0\} \setminus \{x | \cos x = 1\} = \{x | \cos x = -1\}$ .

5)  $\text{ДЕТИ} \setminus \text{МАЛЬЧИКИ} = \text{ДЕВОЧКИ}$ .

Введем еще одно вспомогательное понятие.

Разность между универсальным множеством и произвольным множеством  $A$  называется дополнением множества  $A$  (обозначение:  $\bar{A}$ ). Иными словами,  $\bar{A} = \{x | x \in U \text{ и}$

$x \notin A$ }, или  $A = U \setminus A$ . Например,  $\overline{\text{МОСКВИЧИ}} = \text{НЕМОСКВИЧИ}$ ;  $\overline{\text{ДЕТИ}} = \text{ВЗРОСЛЫЕ}$ ;  $\{x|x < a\} = \{x|x \geq a\}$ .

Закончим рассказ об основных операциях над множествами серией контрольных вопросов.

8. Какие из следующих соотношений справедливы:

- а)  $A \cup \emptyset = A$ ; б)  $A \cap \emptyset = \emptyset$ ; в)  $A \cap U = A$ ; г)  $A \cap \emptyset = \emptyset$ ; д)  $A \cup \overline{A} = U$ ;  
е)  $A \setminus A = \emptyset$ ?

9. Как можно назвать объединение множеств:  $\{,\} \cup \{;\} \cup \{.\} \cup \{|\} \cup \{?\} \cup \{...\} \cup \{:\} \cup \{-\}$ ?

10. Решить систему уравнений (найти неизвестное  $X$ , где  $X$  — множество):

$$\begin{cases} \{1; 2; 3;\} \setminus x = \{1\}. \\ \{x \cup \{3\} = \{2, 3, 4\}. \end{cases}$$

11. Чему равно пересечение множества физиков нашей страны с множеством лауреатов Нобелевской премии?

## 1.В. Кортеж

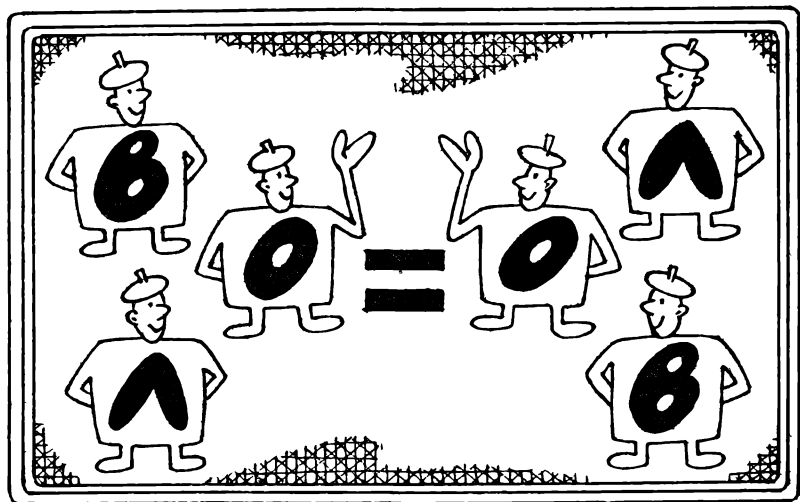
**Понятие кортежа.** Если с термином «множество» вам, видимо, приходилось сталкиваться на уроках в школе, то математический термин «кортеж» вам, вероятно, неизвестен. Между тем понятие кортежа имеет фундаментальное значение для любой формальной теории. Почему это так, надеемся, вы поймете сами.

Кортеж будем считать понятием неопределяемым. Скажем лишь, что аналогом кортежа является известное вам из курса математики понятие последовательности. Мы будем рассматривать только кортежи, состоящие из конечного числа элементов. Кортеж, содержащий  $n$  элементов, называется *n-элементным* или *n-местным*.

Кортеж, как и множество, может состоять из элементов любой природы. Однако кортеж отличается от множества следующими свойствами: 1) его элементы линейно упорядочены (вспомните: первый член последовательности, второй член и т.д.); 2) его элементы могут совпадать.

Кортеж, состоящий из  $n$  элементов, обычно обозначают так:  $\langle a_1; a_2; \dots; a_n \rangle$ .

В жизни кортежи встречаются на каждом шагу. Прежде всего всякое слово (и даже текст!) может рассматриваться как кортеж конечной длины. Так, слово *слово* — это 5-местный кортеж  $\langle c; л; о; в; о \rangle$ ; слово *вол* — 3-местный кортеж  $\langle в; о; л \rangle$ , а слово *лов* — другой 3-местный кортеж  $\langle л; о; в \rangle$  (очевидно, что кортежи  $\langle a_1, \dots, a_n \rangle$  и  $\langle a'_1, \dots, a'_n \rangle$  совпадают в том и только в том случае, если  $a_i = a'_i$  для любого  $i$  от 1 до  $n$ ). Любой список — это тоже кортеж; элементы списка часто упорядочиваются по алфавиту но не обязательно (пример — список



дежурных по классу). Шахматная партия — это кортеж ходов, номер телефона — кортеж цифр, свадебный кортеж — кортеж соответствующих автомобилей, программа на языке Бей-сик — кортеж команд.

Кортеж, как и множество, может не содержать ни одного элемента. Такой кортеж называется п у с т ы м, или н у л ь м е с т н ы м. Он обозначается  $\Lambda$ .

**Действие над кортежами.** Для кортежей определяется особое действие, которое позволяет по двум кортежам построить третий. Это действие носит название к о н к а т е н а ц и я (или сцепление, приписывание).

Конкатенацией двух кортежей:  $m$ -местного кортежа  $A$  и  $n$ -местного кортежа  $B$  — называется  $(m+n)$ -местный кортеж, такой, что его первые  $m$  элементов совпадают с элементами кортежа  $A$ , а следующие  $n$  элементов совпадают с элементами кортежа  $B$ . Обозначение:  $A \oplus B$ . Ясно, что не всегда  $A \oplus B$  равно  $B \oplus A$ . Естественно считать, что  $A \oplus \Lambda = \Lambda \oplus A = A$  ( $A$  — произвольный кортеж).

Любопытный пример применения операции  $\oplus$  в словесном творчестве дает составление шарад. «Целое» (разгадка) представляет собою конкатенацию составных частей, например:  $ком \oplus пас = компас$ . В этой схеме шарады зашифровываются части  $ком$  и  $пас$ , и слушающему предлагается восстановить результат  $компас$ .

Для некоторых кортежей в математике приняты специальные названия. Двухэлементный кортеж называется «парой», трехэлементный — «тройкой» и т. п.

## 1.С. Отношение

**Формальное понятие отношения.** В жизни мы часто встречаемся с различными отношениями, определяющими взаимную связь между какими-либо объектами (предметами, множествами, величинами, понятиями, событиями и т. д.). Например, следующие предложения выражают отношения между некоторыми объектами: *Леня — родной брат Маши, Коля — сосед Пети, Ростов южнее Мурманска, Виктор Николаевич старше Николая Викторовича, 5 больше, чем 4, Слово кот и слово ток состоят из одних и тех же звуков* и т. д.

Можно заметить, что каждое из отношений («быть родным братом», «быть соседом», «быть южнее» и др.) связывает два объекта, которые принадлежат некоторому одному множеству (сына и дочь общих родителей, жителей одного дома или рядом стоящих домов, городов и т. д.). Вообще говорить об отношении можно лишь тогда, когда указано множество элементов, на котором это отношение рассматривается. Данное обстоятельство необходимо учитывать при переходе от интуитивного и обыденного понятия отношения к его формальному аналогу. Формальное определение понятия отношения строится на основе понятия «пары», а именно: отношением на множестве  $M$  называется множество пар, таких, что элементы этих пар принадлежат множеству  $M$ . Если пара  $\langle x, y \rangle$  входит в отношение  $R$  (как в множество пар), то принято говорить, что  $x$  находится в отношении  $R$  к  $y$  (или что  $x$  и  $y$  связаны отношением  $R$ ). Обозначение:  $R(x, y)$ .

Приведем некоторые примеры отношений.

1) Пусть  $A = \{1; 2; 3\}$ . Отношение  $\{\langle 2; 1 \rangle; \langle 3; 1 \rangle; \langle 3; 2 \rangle\}$  называется «больше» (на множестве  $A$ ): 2 больше 1; 3 больше 1; 3 больше 2.

2) Пусть  $C$  — множество согласных звуков русского языка. Тогда отношение  $\{\langle [б]; [п] \rangle; \langle [б']; [п'] \rangle; \langle [п]; [б] \rangle; \langle [п']; [б'] \rangle; \langle [д]; [т] \rangle; \langle [д']; [т'] \rangle; \langle [т]; [д] \rangle; \langle [т']; [д'] \rangle; \dots\}$  называется отношением парности по звонкости-глухости на  $C$ .

3. Отношение  $\{\langle \text{Онегин}; \text{Онегин} \rangle; \langle \text{Онегин}; \text{Ленский} \rangle; \langle \text{Ленский}; \text{Онегин} \rangle; \langle \text{Онегин}; \text{Татьяна Ларина} \rangle; \dots; \langle \text{Чичиков}; \text{Чичиков} \rangle; \langle \text{Чичиков}; \text{Манилов} \rangle; \dots\}$  на множестве литературных героев — это отношение «быть героями одного произведения».

Основные способы задания отношений — такие же, как и для любого множества: отношение может быть задано перечислением элементов или указанием характеристического свойства пар.

**Свойства отношений.** Возьмем два привычных нам отношения на множестве всех действительных чисел: «меньше» ( $<$ )

и «меньше или равно» ( $\leq$ ). Чем они отличаются друг от друга? Известно, как трудно бывает сразу понять утверждение вида «верно, что  $2 \leq 2$ ». Истинность последнего утверждения обусловлена тем, что в отношении «меньше или равно» входит пара  $\langle 2; 2 \rangle$ . По аналогичным причинам верны утверждения  $3 \leq 3$ ;  $4 \leq 4$ ;  $0,5 \leq 0,5$  и т. д. В то же время высказывание  $2 < 2$  ложно, поскольку пара  $\langle 2; 2 \rangle$  не принадлежит отношению «меньше». Мы видим, что эти два отношения отличаются на пары вида  $\langle a; a \rangle$ .

Те отношения на множестве  $M$ , которые (наряду с другими парами) содержат все пары вида  $\langle a; a \rangle$ , где  $a$  — произвольный элемент из  $M$ , называются рефлексивными. Иными словами, отношение  $R$  на множестве  $M$  рефлексивно тогда и только тогда, когда  $R(x, x)$  для любого  $x$  из  $M$ . В частности, к рефлексивным относится отношение «меньше или равно» на множестве натуральных чисел. Кроме того, рефлексивными являются такие отношения, как отношение тождества, или равенства, на произвольном множестве, отношение «принадлежать к одной части речи» на множестве слов, отношение «жить в одном доме», отношение «не старше» на множестве людей и т. п.

Важную роль в лингвистике играют также симметричные отношения. Отношение  $R$  на множестве  $M$  называется симметричным, если для любых двух элементов  $x$  и  $y$ , принадлежащих  $M$ , из  $R(x, y)$  следует  $R(y, x)$ . Иными словами, пусть пара  $\langle x; y \rangle$  входит в  $R$ . Тогда если  $R$  симметрично, то пара  $\langle y; x \rangle$  непременно тоже входит в  $R$ .

Приведем примеры. Симметричными являются отношение равенства (на множестве действительных чисел), отношение «дружить» (на множестве людей), отношение параллельности (на множестве прямых), отношения синонимии и антонимии (на множестве слов данного языка). В то же время отношения «меньше», «меньше или равно» (на множестве чисел) и «не старше» (на множестве людей) симметричными не являются.

Еще одно важное свойство отношений — свойство транзитивности. Отношение  $R$  на множестве  $M$  является транзитивным, если для любых трех элементов  $x$ ,  $y$  и  $z$ , принадлежащих  $M$ , из  $R(x, y)$  и  $R(y, z)$  следует  $R(x, z)$ . Иными словами, пусть отношением  $R$  связаны, с одной стороны, элементы  $x$  и  $y$ , а с другой —  $y$  и  $z$  (в указанном порядке). Тогда, если  $R$  транзитивно, то в него непременно входит пара  $\langle x, z \rangle$ .

Транзитивными являются неоднократно уже упоминавшиеся нами отношения «меньше», «меньше или равно», отношения равенства, параллельности, отношение «быть под-

множеством». Не являются транзитивными отношения «быть знакомым», «быть отцом» (на множестве людей), отношение «сниматься в одном кинофильме», отношение «пересекаться» на множестве прямых и др.

Если отношение  $R$  одновременно рефлексивно, симметрично и транзитивно, то говорят, что  $R$  — отношение эквивалентности. Отношения эквивалентности бывают самые разнообразные. Это такие отношения, как равенство, равносильность (на множестве уравнений), принадлежность к одной части речи (на множестве слов), отношение «быть ровесниками» и т. п.

Подчеркнем, что эквивалентность — это название свойства отношения, а не какого-то конкретного отношения. Очевидно, что на одном и том же множестве можно задать несколько разных отношений эквивалентности.

Понятие эквивалентности очень важно для построения правильных классификаций. Правильная классификация, или разбиение, элементов некоторого классифицируемого множества  $A$  предполагает, что получающиеся в результате подмножества, или классы, обладают следующими свойствами:

- 1) любой элемент множества  $A$  входит в тот или иной класс, т.е. объединение классов равно  $A$ ;
- 2) никакие два класса не имеют общих элементов (пересечение любых двух классов равно  $\emptyset$ ).

Связь между отношениями эквивалентности и правильными классификациями отражена в следующей важной теореме, известной в математике под названием «теорема о разбиении»:

- 1) свойство «принадлежать одному классу разбиения множества  $M$ » задает отношение эквивалентности;
- 2) каково бы ни было отношение эквивалентности на множестве  $M$ , можно построить такое разбиение множества  $M$ , что любые два элемента, связанные данным отношением эквивалентности, попадут в один класс, а не связанные данным отношением — окажутся в разных классах.

Доказательство теоремы о разбиении увело бы нас слишком далеко в сторону. Желаящие могут познакомиться с доказательством в книгах: *Шиханович Ю. А.* Введение в современную математику. — М., 1965; *Шрейдер Ю. А.* Равенство, сходство, порядок. — М., 1970.

Значение этой теоремы выходит далеко за пределы математики. Она используется в любой науке, связанной с классификацией объектов.

Закончим наш рассказ про теорему о разбиении следующей задачей.





12\*. Доказать или опровергнуть следующее утверждение: участников городской олимпиады по языкознанию и математике всегда можно рассадить по разным аудиториям так, чтобы выполнялись два условия:

1) если фамилии двух участников содержат хотя бы одну общую букву, то эти два участника попадают в одну аудиторию;

2) фамилии любых двух участников, попавших в одну аудиторию, содержат хотя бы одну общую букву.

## 1. Д. Знаки и знаковые системы

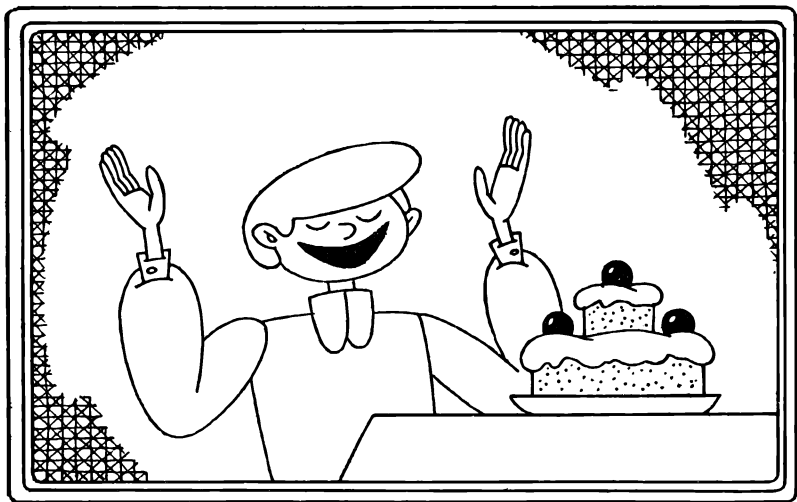
С самого рождения человека окружают различного рода знаки. Плач ребенка — это нередко знак того, что он хочет есть или у него что-то болит. Телефонный звонок — знак того, что кто-то набрал ваш номер и надо снять трубку. Красный свет светофора — знак того, что движение запрещено. Когда мы говорим или рассказываем о каких-то вещах, то обычно не показываем их собеседнику, а пользуемся названиями вещей — словами. В этом смысле можно сказать, что слово служит знаком вещи.

Итак, мы видим, что все объекты и явления реального мира можно разделить на два класса: передающие информацию о других объектах или явлениях (бой часов, телефонный звонок, слово *торт*, сигнал светофора и др.) и не передающие такой информации (часы, светофор, торт, движение автомобиля). Объекты первого класса принято называть **знаками**.

Граница между знаками и не знаками весьма условна. Слово *торт* является знаком весьма вкусного объекта, обозначаемого этим словом; торт на столе, в свою очередь, может быть знаком того, что в доме праздник. Мы видим, что любой объект становится знаком, если используется как знак, т.е. передает информацию о другом объекте или явлении.

Одному объекту могут соответствовать разные знаки. Например, торт по-русски называется *торт*, по-английски — *cake*, по-французски — *tarte*, по-немецки — *Torte*. Слова *Пушкин*, *автор «Евгения Онегина»* и *поэт, убитый Дантесом на дуэли*, обозначают одно и то же лицо. Выражения  $2 + 3$  и  $4 + 1$  обозначают одно и то же число. Наоборот, один знак может обозначать в разных ситуациях разные объекты. Например, поднятая рука ученика может означать, что он хочет ответить на вопрос учителя или что он просит разрешения выйти. Как мы уже видели, слово *кортеж* является знаком двух вещей: математического понятия и вереницы экипажей.

В науке принято всякий **знак** рассматривать как **кортеж**, состоящий из двух элементов: означающего и означае-



мого. О з н а ч а ю щ е е — это материальная оболочка знака: звучание слова, звук гонга, исходящий от светофора свет. О з н а ч а е м о е — это смысл знака: значение слова, информация, которую передает звук гонга, информация о том, что движение в данном направлении разрешено или запрещено. Именно благодаря смыслу объект, не являющийся знаком, превращается в знак.

Означающее знака чаще всего бывает оптическим, т.е. воспринимаемым зрением (сигнал светофора, поднятая рука, морская сигнализация флажками), или акустическим, т.е. воспринимаемым на слух (гонг, телефонный звонок, плач ребенка).

Самый важный вид знака — слово — может иметь означающее как оптической, так и акустической природы. Написанное слово имеет оптическое означающее (его материальная оболочка — это последовательность букв), а произнесенное слово обладает акустическим означающим (его материальная оболочка — это последовательность звуков).

Бывают ситуации, когда означающее материально не выражено; в этом случае само отсутствие материального выражения несет какой-то смысл. Вспомним такие фразы, как «Молчание — знак согласия», «Молчание красноречивей всяких слов» и т. п. Знаки с материально не выраженным означающим называются н у л е в ы м и. Более подробно о некоторых нулевых языковых знаках будет рассказано в главе 4.

Знаки, относящиеся к какой-либо одной области и

связанные друг с другом, образуют знаковую систему. Приведем примеры знаковых систем.

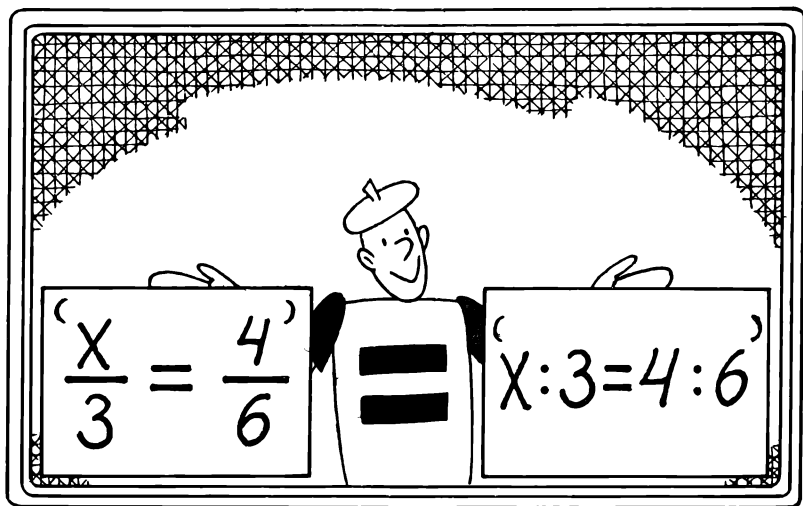
Знаковой является уже упомянутая нами система, состоящая из трех знаков, — сигналы светофора. Также является знаковой система этикета: означающими в ней являются жесты, движения и устойчивые выражения — словесные клише (например, рукопожатие, поклон, звучание произнесенного слова *спасибо* и т. п.), а означаемым — та информация, которую эти жесты, движения и слова передают, т. е. выражение приветствия, уважения, благодарности и т. п.

Язык любой науки представляет собой сложную систему знаков. Можно даже сказать, что изучить язык науки значит овладеть соответствующей знаковой системой.

Наконец, любой естественный язык тоже является системой знаков, причем универсальной, т. е. применимой в любой области. При помощи естественного языка — например русского — можно передать любую нужную информацию.

Общим свойством всех знаковых систем является наличие в них исходных, или элементарных, знаков. Кроме того, многие из знаковых систем содержат и неэлементарные (сложные) знаки, причем эти последние строятся из элементарных по определенным правилам. Так, в языке арифметики сложным знаком является выражение  $2+3$ . Оно построено из трех элементарных знаков: 2, + и 3. В русском языке (в его письменной форме) сложный знак ?! построен из двух знаков.

Из сложных знаков образуются еще более сложные



знаки, и т. д. Иначе говоря, сложный знак состоит из других, более простых знаков. Например, предложение состоит из слов, а текст — из предложений.

Наличие сложных знаков позволяет при помощи ограниченного числа элементарных знаков передавать бесконечное множество разнообразных смыслов.

И простые, и сложные знаки могут, помимо обычного, иметь так называемое **автоимное** употребление. Употребление знака называется **автоимным**, если обозначаемый им объект — сам этот знак. Например, в названии повести М. Алексеева «Хлеб — имя существительное» *хлеб* обозначает не пищевой продукт, а именно слово: ведь продукт не может быть существительным.

Между знаками устанавливаются некоторые важные отношения. Из них мы рассмотрим два: отношение абсолютной синонимии и отношение омонимии.

Два знака являются абсолютно **синонимичными**, если они имеют при разных означающих тождественные означаемые. Два знака называются **омонимичными**, если они имеют при разных означаемых тождественные означающие. Таким образом, абсолютно синонимичные знаки имеют вид  $\langle a, \alpha \rangle$  и  $\langle b, \alpha \rangle$ , а омонимичные —  $\langle a, \alpha \rangle$  и  $\langle a, \beta \rangle$ . Например, в знаковой системе арифметики знаки умножения  $\times$  и  $\cdot$  находятся в отношении абсолютной синонимии. В знаковой системе математики в отношении омонимии находятся знаки  $\Delta$  (как обозначение треугольника) и  $\Delta$  (как обозначение абсолютной погрешности).

Знаки, находящиеся в отношении абсолютной синонимии, как правило, могут быть заменены один на другой в составе сложного знака без изменения его смысла. Например,

записи пропорции  $\frac{x}{3} = \frac{4}{6}$  и  $x : 3 = 4 : 6$  имеют один и тот же

смысл, поскольку знаки — (черта дроби) и  $:$  (знак деления) абсолютно синонимичны. Однако при автоимном употреблении знака замена его на абсолютно синонимичный невозможна. Так, в предложении  $\frac{4}{6}$  *это правильная*

*дробь* нельзя заменить  $\frac{4}{6}$  на знак  $4 : 6$ .



13. Доказать, что отношения абсолютной синонимии и омонимии являются отношениями эквивалентности на множестве знаков (в заданной знаковой системе).

## 2. ГРАММАТИКА РАССМАТРИВАЕМЫХ МЕТАЯЗЫКОВ

Как мы уже говорили, назначение всякой грамматики состоит в том, чтобы задавать правила построения предложе-

ний данного языка. Грамматика каждого из используемых нами метаязыков — математического и алгоритмического — устроена по-своему. Основное отличие их состоит в том, что одна из них, так сказать, утвердительная, а другая — повелительная. Это означает, что для математического языка характерны повествовательные предложения, а алгоритмическому языку в большей степени свойственны побудительные предложения.

Укажем сначала основные типы текстов математического языка. Во-первых, это тексты-рассуждения, или умозаключения. Во-вторых, это определения (о строении определений подробнее будет рассказано в главе 4). Наконец, это аксиомы, теоремы, задачи, упражнения и еще многое другое. Конечно, каждый из этих типов текстов обладает своими особенностями, но здесь мы на них останавливаться не будем.

Переходя к описанию грамматики алгоритмического языка, отметим, что основной его единицей является команда, а более крупные тексты — алгоритмы — строятся из команд. Команда представляет собою побудительное предложение, реакцией исполнителя (машины или человека) на которое должно явиться выполнение соответствующего требования. Общаться на алгоритмическом языке довольно сложно: даже самое простое сообщение приходится подавать в форме приказа, а вопрос — в виде команды «Сообщи...».

Команды в алгоритме соединяются особыми знаками-связями. Описание этих связей составляет отдельный фрагмент грамматики алгоритмического языка. Здесь мы остановимся только на основных блоках, из которых состоит алгоритм. Эти блоки таковы: описание исходных данных, основной каркас алгоритма (последовательность выполняемых команд) и анализ «аномалий», или «вырожденных» случаев, когда действие алгоритма отклоняется от основного пути. Принцип действия каждого из этих блоков мы продемонстрируем на примере работы конкретных лингвистических алгоритмов.

Глава вторая

**Звуки и буквы.  
Формальный  
аспект**



[П'АТ' РОС]



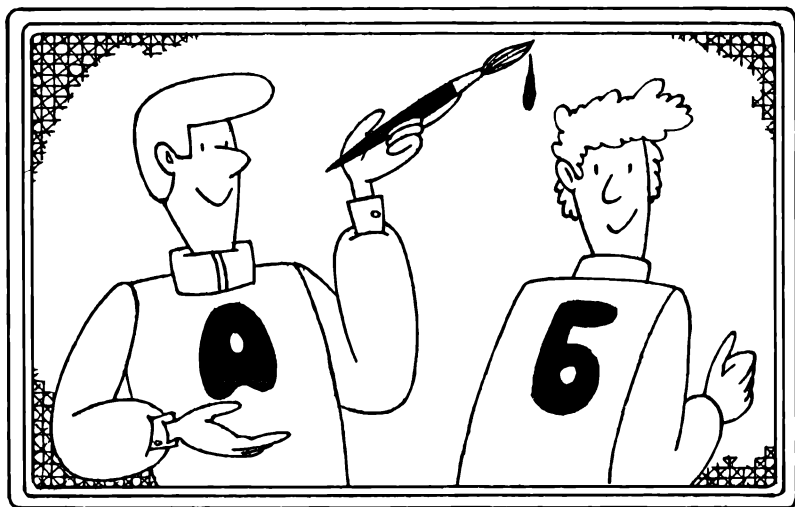
## І. ФОНЕТИКА И ГРАФИКА

### 1. ОБЩЕЕ ПОНЯТИЕ О КЛАССИФИКАЦИИ

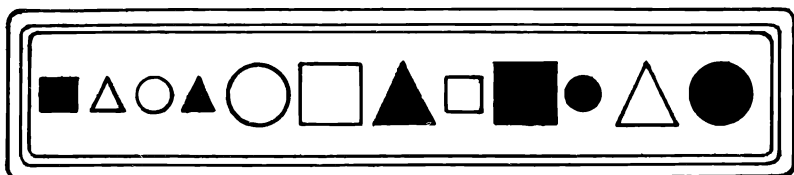
**Д**авайте мысленно вернемся на восемь—десять лет назад. Теплый августовский день, и мама говорит вам: «Сегодня мы идем записываться в школу». Вы приходите туда и слышите, что с 1-го сентября будете учиться в 1 «А» классе. Придя же домой, узнаете, что соседа записали в 1 «Б», т.е. вы с ним попали в разные классы. Произошел простейший акт классификации. Все первоклассники в школе были разбиты на несколько параллельных классов. С классификациями вы сталкивались и до и после этого. Без умения классифицировать различные объекты людям трудно было бы ориентироваться в жизни. Они не могли бы описывать, например, чем хорошие квартиры отличаются от плохих, не могли бы выделять разные виды животных и с трудом находили бы дорогу к незнакомому дому, потому что не могли бы указать признаки, отличающие одну улицу от другой.

#### 1. А. Правильная классификация

**Интуитивное представление о хорошей классификации.** Вообще важным способом изучения какого-нибудь множества объектов является построение их правильной классификации — разбиения.



Между тем даже правильные классификации бывают интуитивно неудовлетворительными. Разберем следующий опыт. Пусть у нас есть двенадцать вырезанных из бумаги геометрических фигур (см. рисунок). Психологи предлагают испытуемым разбить их на группы. Были получены такие результаты (см. рисунок на с.30).



На основании этих результатов психологи сделали для себя выводы: испытуемые *А*, *Б* и *В* умеют классифицировать, а *Г* и *Д* делают это плохо. И мы бы с учеными, по-видимому, согласились. Почему? Видимо, по той причине, что у нас есть внутреннее, интуитивное представление о том, что такое хорошая классификация и что такое классификация плохая, хотя дать формальное объяснение различий между хорошими и плохими классификациями — дело очень трудное. Существует даже целая научная область, решающая подобные проблемы, — теория классификаций.

**Классификационные признаки.** Классификация как составная часть входит в любую науку: классификация объектов является одним из важнейших принципов построения любой научной дисциплины. При этом за основу берутся те признаки, которые отличают один объект от другого. Такие признаки в теории классификаций принято называть дифференциальными, или классификационными, причем хорошую классификацию создают хорошие дифференциальные признаки, т. е. те признаки, которые приложимы именно к данным объектам. Например, объекты можно отличать по цвету, форме, вкусу и т. д. Так, деля предметы на классы по их цвету, мы получим класс белых предметов, класс черных предметов, класс красных предметов и т. д.; деля глаголы по признаку «вид», мы получаем два класса: глаголы совершенного вида и глаголы несовершенного вида.

Условимся говорить, что каждый дифференциальный признак обладает множеством значений. Например, значения признака «цвет» — это «белый», «зеленый», «серый» и т. д.,



**A**



**Б**



**В**



**Г**



**Д**



а значения признака «вкус» — «сладкий», «солёный», «кислый», «горький».

Хотя название признака нередко совпадает с названием одного из его значений, важно не смешивать их в научной практике. Например, признак «четность», определенный на множестве натуральных чисел, принимает два значения: «четность» и «нечетность», при этом название признака «четность» — это не то же самое, что «четность» как имя значения этого же признака.

Как правило, классификации объектов учитывают не один, а сразу несколько дифференциальных признаков. Пирожки бывают жареные и печеные; купленные с лотка и домашние; с повидлом, с яблоками, с мясом, с капустой, с рисом и т. д. Среди фильмов выделяются художественные и документальные, детские и взрослые, односерийные и многосерийные и т. п. Функции подразделяются на непрерывные (на некотором множестве  $M$ ) и разрывные, периодические на  $M$  и непериодические, выпуклые на  $M$  и вогнутые и т. д. Предложения классифицируются по цели высказывания, по интонации и по ряду других синтаксических признаков (выделяют предложения простые и сложные, распространенные и нераспространенные и т. д.).

## 1.В. Виды классификаций

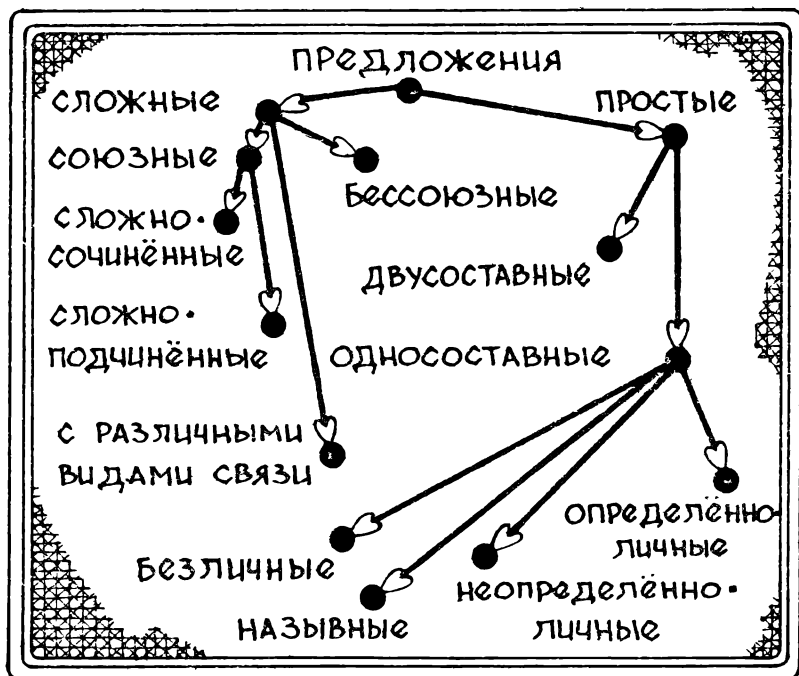
В науке из всех возможных типов классификаций обычно выделяются два вида классификаций объектов по нескольким признакам: универсальная и древовидная. Определить вид классификации можно лишь тогда, когда известны все положенные в ее основу дифференциальные признаки.

**Универсальная классификация.** При универсальной классификации каждому объекту приписывается одно из возможных значений каждого из выбранных дифференциальных признаков. Например, классификация предложений по двум дифференциальным признакам — по цели высказывания и интонации — носит универсальный характер. В самом деле, любому предложению приписывается одно из трех значений признака «цель высказывания» (повествовательное, вопросительное, побудительное) и одно из двух значений признака «интонация» (восклицательная, невосклицательная). Еще один пример универсальной классификации — это классификация причастий. Каждому причастию приписывается значение признака времени (настоящее, прошедшее) и залога (действительный, страдательный). Тем самым выделяется четыре класса причастий: действительное настоящего времени (*читающий*), страдательное настоящего времени

(*читаемый*), действительное прошедшего времени (*читавший*) и страдательное прошедшего времени (*читанный*).

**Древовидная классификация.** Эта классификация называется так потому, что графически она может быть представлена в виде «дерева».

На рисунке представлена древовидная классификация предложений. Сначала все предложения в соответствии с признаком «число предикативных частей» делятся на два класса: простые и сложные. Для каждого из полученных классов имеется свой дифференциальный признак, по которому осуществляется дальнейшее деление. Так, простые предложения делятся по числу главных членов на односоставные и двусоставные, а сложные по типу связи между составными частями — на союзные, бессоюзные и предложения с различными видами связи. Для простых односоставных предложений берется признак «грамматическая характеристика главного члена», в соответствии с которым осуществляется дальнейшее деление; классификация сложных союзных предложений, в свою очередь, основана на признаке «тип союзной связи». Аналогичным образом строится и любая другая древовидная классификация. «Дерево» как формальный объект удобно тем,



что по нему легко вычитываются формальные характеристики классификации, а именно — какие дифференциальные признаки положены в основу классификации и какие значения они принимают.

**Универсально-древовидная классификация.** Поскольку обычно в науке приходится классифицировать объекты крайне сложной природы, получаемые классификации чаще всего имеют смешанный характер. Исследователь может сначала пойти по одному пути и начать, например, с древовидной классификации, а затем на каком-то шаге счесть более удобным переход к универсальной классификации. В результате он получает смешанную, универсально-древовидную, классификацию.

## 2. КЛАССИФИКАЦИЯ ЗВУКОВ

### 2. А. Объекты и признаки классификации

На примере классификации звуков речи мы покажем, как выглядит универсально-древовидная классификация. Классификация звуков интересна тем, что здесь происходит стыковка языкознания, с одной стороны, с акустикой, а с другой — с анатомией и физиологией человека. Действительно, в основу этой классификации кладутся признаки, активно использующиеся в названных областях знания.

**Гласные и согласные.** Сначала все звуки речи делятся на гласные и согласные. Какой же признак положен в основу такого деления? Для ответа на этот вопрос лингвистам пришлось изучить целый раздел анатомии и физиологии — строение и работу голосового тракта человека. Хорошо известно, что звуки — это колебания воздуха. Если колебания происходят с постоянной частотой (как колебания струн при игре на гитаре), звук называется **т о н о м**; в противном случае звук носит название **ш у м**. Этим акустическим различиям можно поставить в соответствие физиологические артикуляционные различия, связанные с работой органов речи у человека. «Тональные» звуки возникают при колебании голосовых связок: в этот момент связки напряжены, и проходящий воздух заставляет их двигаться с постоянной частотой. «Шум» возникает, когда струя воздуха преодолевает препятствие в полости рта: сомкнутые губы, язык, прижатый к верхним зубам или нёбу, и т. д. Легко видеть, что звуки речи могут состоять только из тона, только из шума или включать в себя и то и другое. Звуки, не включающие шум (чисто «тональные»), носят название **г л а с н ы х**; звуки, включающие шум, называются **с о г л а с н ы м и**.

**Согласные глухие и звонкие.** Согласные бывают двух типов: произносимые без участия голоса (т. е. состоящие только из шума) и предполагающие участие голоса (т. е. включающие как шум, так и тон). Первые называются **глухими**, вторые — **звонкими**.

**Согласные мягкие и твердые.** Для классификации согласных русского языка принято использовать еще один дифференциальный признак. Этот признак связан с расположением средней части языка в момент произнесения звука. Если средняя часть языка приподнята к нёбу, то согласный звучит «мягко» и называется **мягким согласным**; если же эта часть не поднята, то согласный звучит «твёрдо» и называется **твёрдым согласным**.

**Гласные ударные и безударные.** Гласные звуки, в свою очередь, делятся на **ударные** и **безударные**. Этому делению соответствуют свои акустические и артикуляционные признаки.

Таким образом, мы набросали фрагмент классификации русских звуков. Она, очевидно, является смешанной, древно-универсальной.

Может возникнуть вопрос: почему в основу классификации положены именно эти, а не какие-то другие признаки? На первый взгляд кажется, что выбор признаков совершенно произволен (они, так сказать, «берутся с неба»). Но на самом деле выбранные признаки далеко не случайны и используются при решении целого ряда практических задач. Так, они «работают», например, в системе орфографических правил: вспомним правила правописания безударных гласных, звонких и глухих согласных, способы обозначения мягкости согласных на письме.

## **2.В. Собственно классификация**

В принципе не исключено, что для других целей более важными окажутся другие дифференциальные признаки. В науке строится классификация согласных с использованием таких признаков, как «место образования звука» и «способ образования». «Место образования» — это то место в полости рта, где создается преграда, вызывающая шум. По месту образования русские согласные делятся на губные ([б], [б'], [п], [п']), губно-зубные ([в], [в'], [ф], [ф']), зубные ([д], [д'], [т], [т'], [з], [з'], [с], [с'], [ц]), передненебные ([ж], [ш], [ш'], [ч']), средненебные ([й']) и задненебные ([г], [г'], [к], [к'], [х], [х']).

Способ образования — это тип преграды, создающей шум. Это может быть щель, через которую воздух проходит с сильным трением ([в], [в'], [ф], [ф'], [з], [з'], [с], [с'], [ж], [ш], [ш'], [х], [х']), смычка, которую воздух с шумом размыкает

([б], [б'], [п], [п'], [д], [д'], [т], [т'], [й'], [г], [г'], [к], [к']), смычка, переходящая в щель ([ц], [ч'])).

В одной из работ известного писателя В.Т.Шаламова\* высказано мнение, что поэтический смысл формируется на основе особых фонетических классов, которые как раз и создаются на основе признаков «место образования» и «способ образования» согласного. По частоте появления соответствующих классов выделяются особые сочетания согласных — особые звуковые повторы, или, как назвал их сам писатель, «опорные трезвучия» стиха. По мысли Шаламова, эту классификацию «не только каждый школьник, но и всякий берущийся за поэтическое перо должен знать лучше таблицы умножения». «Опорные трезвучия» не только «необычайно расширяют видимый и невидимый мир поэта, но и ограничивают его, ставя какие-то преграды, рамки русской грамматики. Они определяют звуко-смысловую тональность стиха подобно основным музыкальным трезвучиям». Законы «трезвучий», по мнению поэта, важнее, чем звукопись. «Когда Блок пишет: *Зашуришали тревожно шелка* — он делает это не затем, чтобы до наших ушей донести шелест шелкового платья, а затем, чтобы укрепить трезвучия, на которых держится стихотворение» (В.Т.Шаламов).

Вообще не существует звуковой организации поэтической речи, которая не была бы основана на той или иной классификации звуков. При этом всякая общенаучная классификация звуков (например, типа той, что приведена выше) может сложным образом переплетаться с индивидуальной поэтической классификацией.

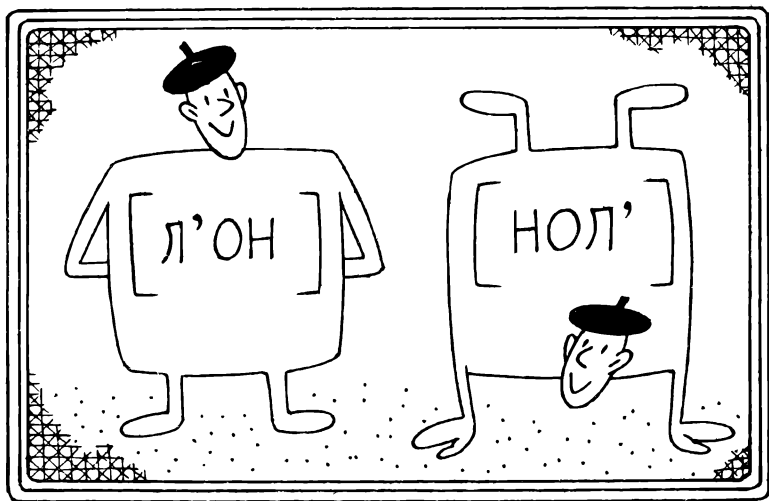
### 3. ФОНЕТИЧЕСКАЯ ТРАНСКРИПЦИЯ

#### 3.А. Соотношение фонетической и буквенной записи

На основе фонетической классификации, как вам уже должно быть известно из уроков русского языка, строится фонетический разбор слова, в частности ф о н е т и ч е с к а я т р а н с к р и п ц и я. При этом если классификация формальная, то она позволяет алгоритмизировать все процессы фонетического разбора и выполнить его на ЭВМ. Такого рода автоматическая обработка русского слова может быть весьма полезна иностранцу, самостоятельно изучающему русский язык. Имея перед собой компьютер, который по запросу выдает чтение любого слова, обучающийся может проверить, освоил ли он правила чтения. Такой же алгоритм применяется

---

\* См.: Шаламов В.Т. Звуковой повтор — поиск смысла (заметки о стиховой гармонии) // Семиотика и информатика. Вып.7. — М., 1976. — С.128—152.



и в так называемых «читающих автоматах», т.е. устройствах, получающих на входе письменный текст и читающих его вслух.

Необходимость создания специального алгоритма перевода буквенной записи слова в фонетическую транскрипцию обусловлена тем, что нет взаимно однозначного соответствия между множеством букв и множеством звуков русского языка. Несоответствие буквенного и звукового облика слова не позволяет осуществлять простой побуквенный перевод. Со звуковой точки зрения, слова *лен* и *ноль* являются «перевертышами»: они состоят из одних и тех же звуков, но идущих друг за другом в разном порядке — [л'он] и [нол']. Между тем в буквенном отношении эти слова вовсе не «перевертыши»: ведь даже количество букв в них различно.

Отсутствие однозначного соответствия между звуками речи и буквами вызвано особенностями русской графики и фонетики.

Во-первых, русская графика носит слоговой характер: звуковое значение букв *я*, *е*, *ё*, *ю* и функции буквы *ь* могут быть определены только в составе слога. Например, буква *я* в словах *яма*, *паяц*, *обезьяна* обозначает два звука [йа], а в слове *пять* — один звук [а] и мягкость предшествующего согласного (аналогично буквы *е*, *ё*, *ю*); буква *ь* в слове *пять* обозначает мягкость согласного [т'], а в слове *обезьяна* — это в первую очередь «разделительный знак»; в слове *паяц* буква *п* обозначает твердый звук [п], а в слове *пять* та же буква обозначает мягкий [п']. Таким образом, как мы видим, алгоритм перевода



требует отдельного блока правил «чтения» букв *я, е, ё, ю* и установления твердости — мягкости согласных.

Вторая особенность уже не графическая, а фонетическая. Она связана с особенностями чтения согласных на конце слова и перед другими согласными. Например, в слове *дуб* пишется *б*, а читается [п], в слове *сбор* пишется *с*, а читается [з]. Тем самым в алгоритме должен быть предусмотрен отдельный блок правил обработки согласных по признаку глухости — звонкости.

Наконец, еще одна особенность чтения русских слов состоит в том, что в русском литературном языке безударные гласные произносятся не так, как ударные. Например, буква *о* в слове *вода* читается иначе, чем та же буква в слове *воды*: в последнем она обозначает звук [о], а в первом — звук, близкий к [а]. Эта особенность также должна быть учтена в алгоритме.

### 3.В. Алгоритм перевода буквенной записи в фонетическую

Теперь можно кратко рассказать идею самого алгоритма. На вход алгоритма поступает слово текста в буквенной записи, но имеющее при себе дополнительное указание — место ударения. Работа алгоритма существенным образом опирается на классификацию символов (к которым относятся как обычные буквы, так и знаки транскрипции). Для удобства мы будем говорить о символах как о звуках: «глухой согласный символ», «гласный символ» и т.д.

Алгоритм перевода удобно разбить на несколько этапов в соответствии с описанными свойствами русской графики и



фонетики. На первом этапе устанавливается чтение букв *е, ё, ю, я* и выявляются мягкие согласные. При этом используется операция подстановки, т.е. одна цепочка символов заменяется на другую. В зависимости от местоположения в слове и характера соседних символов обрабатываемые гласные заменяются на двухсимвольные сочетания (*йэ, йо, йу, йа*) либо на одиночные символы (*э, о, у, а*). В последнем случае предыдущий согласный символ (если это не *(ж), (ш)*) получает знак мягкости. Знак мягкости приписывается и согласным (кроме *(ж), (ш)*), стоящим перед *ь*, а также на этом этапе работы алгоритма устраняются непроезжие согласные и проводится еще ряд вспомогательных операций.

На втором этапе ищут глухие и звонкие согласные, и в тех случаях, когда происходит оглушение или озвончение, символ заменяется на парный по признаку глухости-звонкости. Этот этап также завершается рядом «мелких» вспомогательных операций (замена *сч* на *щ* и т.п.).

На третьем этапе выявляются особенности произношения безударных гласных. Именно здесь-то и используется предварительная информация о месте ударения в слове.

На этом действие алгоритма перевода заканчивается.

Таким образом, наш алгоритм преобразует слово *корова* в *карова*, слово *сбить* в *зб'ит'* и т.д. Действие этого алгоритма не может охватывать исключения, т.е. обрабатывать те слова русского языка, произношение которых противоречит общим правилам чтения (*что, сегодня, ничего* и ряд других). Тем самым исключения из правил чтения остаются исключениями и для алгоритма. Предлагаем задачи.



14. Построить классификацию многоугольников, изучаемых в школе на уроках геометрии. Указать дифференциальные классификационные признаки и их значения. К какому типу относится ваша классификация (древовидная, универсальная, смешанная)?

15. Даны русские слова: *конь, сел, луч, стол, год, шкура, люк, шел, край, сдал, пух*. Определить для каждого из этих слов, сможет ли иностранец, не знающий ни одного русского слова, но знающий правила чтения русских слов, однозначно определить по написанию, как звучит это слово по-русски.

## II. ГРАФИКА И ОРФОГРАФИЯ

### 1. ПРОБЛЕМА ПЕРЕХОДА ОТ ЗВУЧАЩЕЙ РЕЧИ К ПИСЬМЕННОЙ

Выше мы рассматривали вопрос о построении алгоритма, позволяющего перейти от текста в обычной буквенной записи к тексту, записанному в виде фонетической транскрипции

(т. е. алгоритма, «читающего» письменный текст). В повседневной жизни нам часто приходится решать обратную задачу: переходить от звучащей речи к письменной. Мы могли бы назвать много примеров практической деятельности, основанной на решении данной задачи. Это письмо под диктовку, работа машинистки, которая печатает с голоса, и вообще любая запись на бумаге произнесенного вслух текста. Хорошо было бы алгоритмизировать большую часть всей этой работы. Тогда можно было бы купить в магазине автоматическое устройство, которое заменило бы человека в этой деятельности.

Вы можете задать вопрос: что же мешает решить данную задачу? Ведь практическая полезность такого автомата очевидна. Оказывается, основная трудность (причем весьма значительная) состоит в том, что теоретически одно и то же звучание можно передавать различными способами, но при этом только одну запись мы готовы считать правильной. Слово [суп] надо писать *суп*, а не «суб», а слово [дуп] — *дуб*, но никак не «дуп». Слово [ш'от] теоретически можно было бы написать «шёд», «щот», «шёт», «шод», «счот», «счёд», «счод», но правильна лишь одна запись — *счёт*. И лишь в очень немногих случаях существует ровно одна теоретически возможная запись: слово [дом] можно без всяких колебаний записать как *дом* и никак иначе.

Таким образом, поставленную задачу удобно разбить на две: 1) получение для данного звучания всех теоретически допустимых вариантов записи слова; 2) последующий выбор из этих вариантов одного правильного. Лингвистическая дисциплина, основывающаяся на правилах первого типа, называется **г р а ф и к о й**, а правила второго типа относятся к области другой дисциплины — **о р ф о г р а ф и и**.

## 2. АЛГОРИТМИЗАЦИЯ ГРАФИЧЕСКИХ И ОРФОГРАФИЧЕСКИХ ПРАВИЛ

Между графикой и орфографией есть принципиальное различие. Правила графики носят универсальный характер: для применения графического правила не нужно никаких сведений о рассматриваемом слове, кроме его звучания. Эти правила равно применимы к известным словам и к впервые услышанным.

Иначе обстоит дело с орфографическими правилами. Они требуют обращения к другим сведениям о слове, помимо собственно звучания: например, правила написания проверяемых безударных гласных предполагают умение подобрать однокоренное слово, а это невозможно сделать, не понимая его значения; для правильного написания непроверяемых

безударных гласных необходимо просто помнить, как пишется слово.

Отсюда ясно, что вся трудность создания автоматической системы, записывающей звучащие тексты, связана с орфографическими правилами. Ведь чтобы построить такие системы, нужно научить ЭВМ не только воспринимать звучащую речь, но и понимать ее смысл.

Напротив, графические правила сравнительно легко поддаются алгоритмизации именно по той причине, что обращения к смыслу не предполагают. На вход соответствующего алгоритма поступает фонетическая транскрипция, на выходе мы должны получить все теоретически возможные написания слова.

При определении написания слова удобно просматривать фонетическую транскрипцию подряд без пропусков слева направо и определять буквенную запись каждого отдельного звука. При этом следует учитывать правила двух видов.

К первому виду относятся правила обозначения мягких согласных на письме, правила обозначения звука [й'] и правила написания гласных звуков [и] и [ы] после шипящих. Все эти правила дают однозначный результат; их применение не требует последующего обращения к орфографическим правилам.

Ко второму виду относятся правила, дающие сразу несколько вариантов буквенного обозначения одного звука. Для выбора единственного правильного написания необходимо применить орфографические правила. Сюда относятся правила, фиксирующие различные способы обозначения безударных гласных, близких к [а] и [и], обозначение звука [о] после шипящих, обозначение глухих согласных звуков на конце слова и другие.

Предлагаем задачи.



16. («Обратная» к задаче 14.) Для каждого из слов, указанных в задаче 14, определить, сможет ли иностранец, не знающий ни одного русского слова, но знающий основные правила русской графики, по звучанию данного слова однозначно определить его написание.

17. Сколько существует теоретически допустимых способов записи слова [раш'от]? Какой из них является правильным?

Глава третья

**Математические  
методы  
в лексикологии  
и лексикографии**



# 1. СЛОВАРИ И ИХ ФОРМАЛЬНОЕ СТРОЕНИЕ

## 1. ВИДЫ СЛОВАРЕЙ

### 1.А. Энциклопедические и лингвистические словари

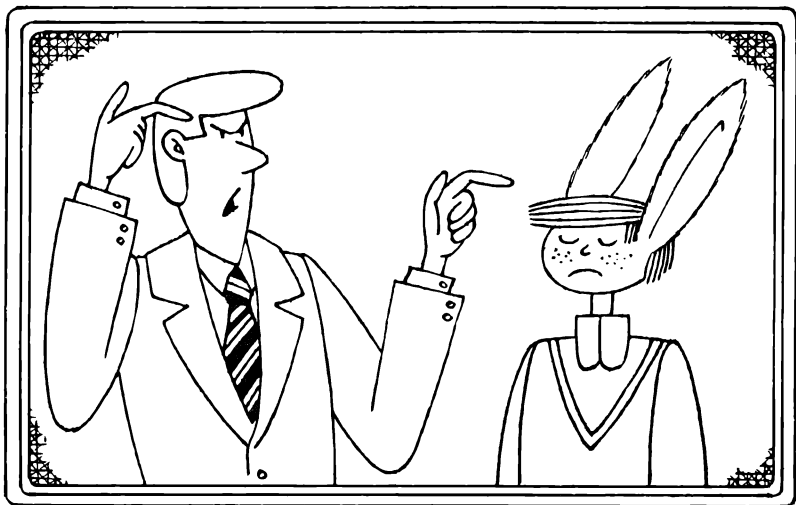
«Приходилось ли тебе пользоваться словарями?» — спросила учительница. «Конечно, — сказала Маша. — Вот недавно мама где-то достала энциклопедический словарь, и я хотела посмотреть, что такое *фривольный*. Но этого слова там не оказалось. Наверное, авторы словаря просто забыли про него». — «Что ж тут странного, что ты его там не нашла? Лучше загляни в словарь иностранных слов. Там-то оно наверняка есть».

В этом диалоге упоминаются два словаря: энциклопедический и лингвистический. Эти два класса — класс энциклопедических и класс лингвистических словарей — составляют основу классификации всех словарей. Признак, различающий энциклопедические и лингвистические словари, удобно описать так: в энциклопедических словарях приводятся сведения об объектах, обозначаемых словами, а в лингвистических — о самих словах. Поэтому в энциклопедических словарях вы не найдете глаголов, наречий, служебных слов, там есть только существительные. Зато среди входящих в словарь существительных много собственных имен: географических названий, имен исторических деятелей, ученых, писателей, художников и др.

Если обратиться к энциклопедической словарной статье «Дарвин», то мы узнаем об обстоятельствах жизни и деятельности ученого, но не найдем сведений, касающихся слова *Дарвин*, например о его принадлежности к именам существительным, об особенностях его склонения и т. п.

Напротив, в лингвистических словарях, как правило, нет информации об объектах, а приводятся разнообразные данные о словах и их поведении в языке. Различие, наверно, станет совсем наглядным, если рассмотреть какое-либо слово, входящее и в энциклопедический, и в лингвистический словарь.

Например, в энциклопедии в словарной статье слова *осел* будет дано описание внешнего вида осла, его повадок, среды обитания, потребляемой пищи и т. п.; в то время как в лингвистическом словаре, где рассказывается о слове *осел*, мы найдем информацию о наличии в слове беглой гласной (ср. *осел* — *осла*), о том, что у этого слова есть переносное значение (глупый или упрямый до глупости человек мужского пола), что существительное *осел* относится к мужскому роду; там же



будут приведены типичные сочетания с этим словом, фразеологизмы, в которые оно входит (ср. *буриданов осел*).

Каждый из вышеназванных больших классов словарей делится на подклассы. Например, энциклопедические словари могут подразделяться по признаку «отрасль знания» (географический энциклопедический словарь, медицинская энциклопедия), по признаку «предназначается для...» («Энциклопедия молодой женщины», «Энциклопедический словарь юного филолога»).

В дальнейшем нас будет интересовать классификация лингвистических словарей. Поэтому рассмотрим сначала ее классификационные признаки.

### **1.В. Виды лингвистических словарей**

Прежде всего словари делятся на одноязычные, двуязычные и многоязычные. Вероятно, чаще всего каждому из вас придется пользоваться двуязычными словарями. Они обладают формальным признаком направленности: русско-китайский словарь не то же самое, что китайско-русский. Двуязычные словари обычно используются для перевода с одного языка на другой. Основная информация в англо-русском словаре об английском слове — это то, как оно переводится на русский язык (или, как говорят, каков его русский «переводной эквивалент»).

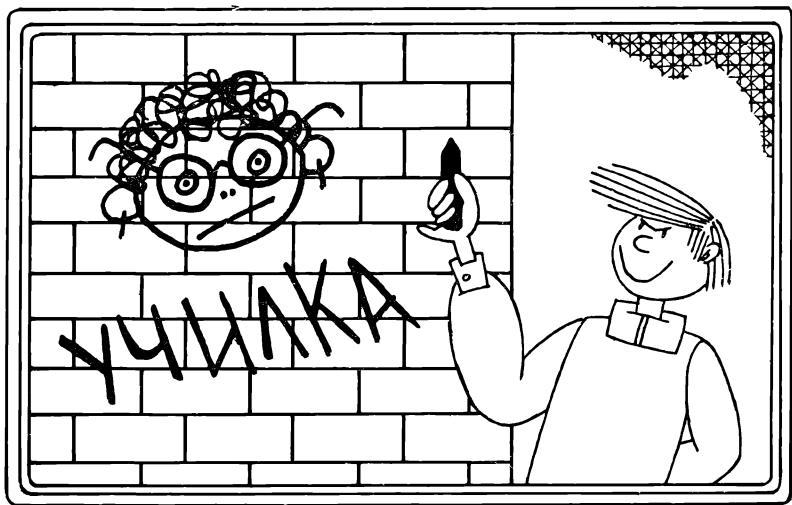
Одноязычный словарь по типу сообщаемых сведений гораздо разнообразнее. И это не удивительно, поскольку он ориентирован на человека, владеющего данным языком.

1. В зависимости от информации о слове различают: орфографический словарь (туда мы заглядываем, когда хотим узнать, как пишется слово), орфоэпический словарь (он отвечает на вопрос, как правильно произнести слово), грамматический словарь (он сообщает, изменяется ли слово и если да, то как), толковый словарь (дает сведения о значении и употреблении слова), этимологический словарь (этот словарь объясняет нам, откуда и как возникло в языке данное слово), словари, в которых указываются «родственники», «друзья» и «враги» данного слова, — омонимы, синонимы, антонимы, частотный словарь (в нем сообщается, насколько употребительно данное слово в текстах) и др.

2. Одноязычные словари можно делить и по другому признаку — по отношению описываемой в нем совокупности слов ко всем словам данного языка. Наряду с общими словарями выделяются словари иностранных слов (они содержат слова, заимствованные из других языков), диалектологические словари (здесь собраны слова, характерные для определенной территории), словари-минимумы (в них входят наиболее употребительные слова языка), словари профессиональной лексики и т. п.

3. Лингвистические словари, как и энциклопедические, различаются и по признаку «предназначается для...». Школьные словари предназначены для школьников, для ребят, обучающихся в училищах, и т. п., существуют словари для иностранцев, журналистов, для работников радио и телевидения.

4. Предназначение словаря — это его внешняя ориентация. Между тем можно провести деление словарей и с точки зрения их внутренней ориентации и, тем самым, внутренней организации. Это известное подразделение словарей на нормативные и описательные. Впервые в русской лексикографической традиции это деление рассмотрел выдающийся русский языковед Л.В.Щерба. К н о р м а т и в н ы м словарям он относил словари, предписывающие, как следует правильно употреблять слова. В них не входят единицы, которые признаются языковым коллективом неправильными, несмотря на то, что часть из них широко употребляется, например такие слова, как *столовка* (вместо *столовая*), *училка* (вместо *учительница*), *директорша*, *читалка* и т. п., формы слова *пиша* (деепричастие), *шофера* (множественное число — вместо *шоферы*) и т. д.; сочетания *играть значение* (вместо *играть роль* или *иметь значение*), *согласно чего* (вместо *согласно чему*) и другие. Цель нормативного словаря — предотвратить «засорение» речи такими выражениями. В о п и с а т е л ь н ы х словарях приводятся все слова (в соответствии с названием



словаря), которые употребляются данным языковым сообществом. В него могут быть, например, допущены все те слова и выражения, которые исключаются нормативным словарем; правда, при всех таких словах должны стоять специальные пометы, указывающие на оценку этих языковых единиц носителями данного языка. Важно подчеркнуть, что такого рода пометы не носят «запретительного» характера (как было бы в нормативном словаре), а объективно констатируют оценку слов языковым сознанием. Таким образом, деление словарей на нормативные и описательные — это различие между словарями, предписывающими, как надо говорить, и словарями, описывающими, как на самом деле говорят. В качестве примера нормативных словарей можно привести орфографические и орфоэпические словари, а в качестве примера описательных — диалектологические словари.

5. По внешнему оформлению словари делятся на «обычные» (как правило, оформляемые в виде книги) и автоматические. Автоматические словари появились сравнительно недавно, и появление их связано с «компьютерной революцией». Автоматические словари — это словари, в которых вся информация записана на машинных носителях (на магнитном диске или магнитной ленте, других запоминающих устройствах). Иными словами, автоматический словарь не существует в отрыве от вычислительной машины.

Как нетрудно увидеть, классификация словарей строится как универсально-древовидная. Она позволяет не только



выделить классы словарей, но и явно указать дифференциальные классификационные признаки.

Теперь, если вновь вернуться к диалогу, приведенному в начале данной главы, то становится вполне понятным совет учительницы. Машу интересовало значение слова *фривольный*, поэтому ей надо было бы обратиться к лингвистическому словарю. А поскольку это слово по происхождению иноязычное, его лучше всего было искать в словаре иностранных слов.

## 2. ТРЕБОВАНИЯ К ВНУТРЕННЕМУ УСТРОЙСТВУ ЛИНГВИСТИЧЕСКИХ СЛОВАРЕЙ

Оставим в стороне энциклопедические словари как не имеющие отношения к изучению языка и обратимся к словарям лингвистическим. Прежде всего важно уяснить, что словарь — это не столько книга для чтения, сколько помощник в нашей интеллектуальной деятельности. Уже одно это обстоятельство накладывает на словари особые требования. Каким бы мы хотели видеть идеальный лингвистический словарь?

Помимо чисто внешних требований (быть сравнительно компактным, приятно оформленным, удобным для чтения), существенным оказывается его внутреннее устройство.

### 2.А. Перечень требований

Основным требованием к внутреннему устройству словаря может быть названа «правильная построенность». Правильно построенным словарем называется словарь, удовлетворяющий следующим требованиям:

1. Он представляет собою непустое множество текстов, из которых один объявляется основным. Правильная построенность словаря — это правильная построенность основного текста.

2. Основной текст правильно построенного словаря — это конечная последовательность мини-текстов, носящих название словарных статей.

3. Каждая словарная статья словаря представляет собою двухместный кортеж вида  $\langle a; b \rangle$ , где  $a$  называется входом в данную словарную статью, а  $b$  — интерпретацией входа. Назначение интерпретации — сообщать определенную информацию о входе.

4. Порядок следования словарных статей в словаре подчиняется некоторому общему принципу, причем этот принцип предполагается читателю заранее известным или извлекается из неосновных текстов словаря.

5. Вход и его интерпретация должны быть связаны некоторым отношением, одним и тем же для всех словарных статей данного словаря. Как и в предыдущем случае, о

содержании этого отношения мы узнаем из неосновных текстов словаря.

6. Среди интерпретаций выделяются те, в которых содержится указание посмотреть другую словарную статью того же словаря (такие интерпретации называются **отсылочными**, а само указание — **отсылкой**). К отсылочным интерпретациям предъявляется еще одно требование. Никакая такая интерпретация не должна отсылать к словарным статьям, которых нет в основном тексте словаря.

7. Содержащаяся в интерпретации информация о входе может быть неоднородной. В правильно построенном словаре информация всегда группируется по **зонам** так, что в каждой зоне информация уже однородна.

8. В правильно построенном словаре порядок зон в составе интерпретации один и тот же для всех словарных статей.

## **2. В. Комментарии к требованиям**

1. Среди неосновных текстов словаря выделяются **название словаря**, **вступительная статья** (предисловие) о целях и задачах словаря, **инструкция** («Как пользоваться словарем») и различные **списки**: **сокращений**, **условных знаков** и другие. Обязательным для любого словаря является наличие названия (заглавия). В заглавии отражается содержательное отношение между входами и интерпретациями; например, в заглавии «Англо-русский словарь» отражено то, что входами являются английские слова, а интерпретациями — их русские переводные эквиваленты. Название «Грамматический словарь русского языка» означает, что в нем интерпретации содержат грамматическую информацию о входах.

2. Обычно в словарях словарные статьи не нумеруются, не именуются, а вычленяются графически. В большинстве случаев каждая следующая статья начинается с красной строки.

3. Составители словарей, как правило, отделяют вход от интерпретации при помощи шрифтовой или какой-то иной разметки. Обычно вход набирается полужирным шрифтом.

4. Словарные статьи в словарях преимущественно упорядочиваются в соответствии с входами. Если входами служат слова, то они, как правило, упорядочиваются по алфавиту.



18. Попробуйте сами придать формальный смысл понятию «упорядоченность слов по алфавиту».

В ряде словарей принят не обычный **алфавитный** порядок расположения слов, а **обратный**. При

обратном порядке сравнение двух слов начинается не с начальных букв, а с конечных и далее продолжается от конца слова к началу. Вот каким может быть начало (начальные входы) такого словаря: *а, ба, баба, жаба, раба, амеба, учеба, ворожба, служба, дружба, тяжба, изба, шайба, колба...* Из двух слов первым идет то, у которого конечная буква ближе к началу алфавита, т. е. к букве *а* (например, *наука* в этом словаре идет раньше, чем *арык*); если последние буквы одинаковы, то упорядочение происходит по предпоследним буквам, и т. д. Таким образом, обратный порядок — это нечто совсем другое, чем перевернутый обычный алфавитный порядок.



19. Даны слова: *щука, поэт, вольт, армия, актер, базальт, асфальт*. Требуется записать их в обычном алфавитном порядке и в обратном.

Существуют и другие способы упорядочения словарных входов. Например, в частотных словарях слова обычно упорядочиваются в соответствии с частотой их употребления: чем выше частота слова, т. е. чем чаще оно употребляется в языке, тем раньше идет в словаре. Один из самых известных частотных словарей русского литературного языка — это словарь Э.А.Штейнфельд, вышедший в Таллинне в 1963 г. Он составлен на базе текстов общим объемом в 400 тысяч слов. Если взять 10 самых часто употребляемых слов, то это — *и, в, на, не, он, я, что, с, этот, быть*. Они и есть первые слова словаря. Самые частые существительные в этом словаре — *день, год, дело, рука*; самые частые глаголы — *сказать, мочь, говорить, знать, идти, хотеть*.

В знаменитом толковом словаре В.И.Даля и в ряде других менее известных словарей используется алфавитно-гнездовой порядок. Этот порядок дает возможность найти в одном месте словаря однокоренные слова. Однокоренные слова во многих словарях имеют сходные интерпретации, сопоставление которых позволяет больше узнать о каждом из этих слов. Но для облегчения поиска рядом помещаются не все однокоренные слова, а только те, которые одинаково начинаются. Такое место в словаре называется алфавитным гнездом. Внутри алфавитного гнезда слова упорядочиваются по алфавиту, а сами алфавитные гнезда следуют друг за другом в соответствии с алфавитным порядком первых слов алфавитных гнезд (отсюда само название «алфавитно-гнездовой»). Возьмем, например, такое множество слов: *влюбиться, влюбленность, любитель, любить, любоваться, любовь, полюбить, самовлюбленность; выстроить, построить, постройка, строитель, строить; выстирать, постирать, стирать, стирка*. В каком порядке шли бы эти слова

при алфавитно-гнездовом расположении? Очевидно, что в таком: *влюбиться*, *влюбленность* (I алфавитное гнездо); *выстирать* (II); *выстроить* (III); *любитель*, *любить*, *любоваться*, *любовь* (IV); *полюбить* (V); *постирать* (VI); *построить*, *постройка* (VII); *самовлюбленность* (VIII); *стирать*, *стирка* (IX); *строитель*, *строить* (X). Иногда в пределах алфавитного гнезда упорядочение происходит не по алфавиту, а некоторым более сложным образом: в паре слов *строить*, *строитель* алфавитный порядок заставляет поставить *строитель* раньше, а реально они могут идти в противоположном порядке в соответствии с интуитивным ощущением большей формальной простоты слова *строить*.

5. Содержание отношения между входом и интерпретацией по большей части раскрывается в названии словаря. В грамматическом словаре интерпретация дает сведения о грамматических свойствах входа, в словаре синонимов — о том, какие у входа имеются синонимы, в орфоэпическом словаре (от греческого *ὀρθός* — правильный и *ῥεῶς* — речь) интерпретация сообщает, как правильно должен произноситься вход, в толковом словаре приводятся толкования (объяснения) значений входов. Однако бывает, что отношение между входами и интерпретациями по заглавию определить трудно. Например, название «Фразеологический словарь» характеризует только входы: оно указывает, что последние представляют собой фразеологические обороты, но ничего не сообщает об интерпретациях. Сведения об интерпретациях извлекаются из других неосновных текстов. В большинстве фразеологических словарей интерпретации представляют собой толкования входов.

6. Теперь речь пойдет об отсылочных интерпретациях. Они позволяют, не нарушая правильной построенности словаря, сократить его объем или установить связь между какими-то его единицами. Существуют стандартные способы введения отсылочных интерпретаций. Это слова *ср.*(сравни), *см.*(смотри), *то же, что...* и другие: например, *калоши* — то же, что *галоши*; *недород* — то же, что *неурожай*; *сладчайший* — см. *сладкий*.

7. В большинстве словарей интерпретация представляет собой текст, разделенный на участки, или так называемые з о н ы. Каждая зона отвечает за определенный тип информации о слове. Так, можно говорить о морфологической зоне, этимологической зоне, зоне иллюстраций, зоне устойчивых сочетаний и т. д. Для отдельных слов те или иные зоны могут оказаться «пустыми»: например, для слова *оазис* зона устойчивых сочетаний пуста. С п о с о б ы у к а з а н и я г р а н и ц м е ж д у з о н а м и в пределах словарной статьи

могут быть разнообразными (само слово *зона* при этом обычно не употребляется). Сюда относятся тип шрифта, цифровая нумерация, особые кавычки и др. Обычно способ выделения зон и способ представления информации в каждой зоне составляет содержание вступительной статьи словаря (руководство «Как пользоваться словарем»).

8. Из множества возможных порядков расположения зон в пределах словарной статьи (в словарях данного типа) обычно один признается более или менее стандартным, и именно его стремятся придерживаться составители соответствующих словарей. Например, в толковых словарях интерпретация начинается с **фонетической зоны**: указываются место ударения в слове и индивидуальные особенности произношения (отклонения от стандартных правил чтения). При этом место ударения обычно помечается особым значком (/) над ударным гласным, и тем самым фонетическая зона графически не отделяется от входа. Далее идет **грамматическая зона**, где сообщается о части речи, к которой принадлежит вход, о его грамматических признаках (род для существительных, вид и управление для глаголов и т. п.), некоторых формах изменения для изменяемых слов. Затем идет **зона специальных помет**, например *разгов.* — разговорное, *устар.* — устаревшее, *книжн.* — книжное, *ласк.* — ласкательное, *груб.* — грубое, *обл.* — областное, *редк.* — редкое и другие. Эти пометы характеризуют сферу употребления слова, его эмоционально-оценочную окраску, частоту употребления и т. п.

Информацию в зонах принято распределять по отдельным значениям слова. Каждому значению присваивается свой номер. Фрагменты интерпретации, соответствующие разным значениям слова, следуют друг за другом в порядке возрастания нумерации значений: сначала идет вся информация, соответствующая первому значению, затем — второму и т. д.

В составе каждого из выделенных значений интерпретация представляется также единообразно. После арабской цифры (номера данного значения) идут грамматические сведения и пометы, характеризующие особенности именно данного значения. Наряду с уже упомянутыми пометами (*разгов.*, *устар.* и т. д.), здесь может ставиться помета *перен.*, означающая, что данное значение возникло из некоторого другого на основе смыслового «переноса». Например, этой пометой снабжается второе значение слова *осел* (глупый или упрямый до глупости человек мужского пола).

Основная зона толкового словаря — это, очевидно, **толкование**. Оно следует за пометами, относящимися к данному значению, или непосредственно за цифрой в случае

отсутствия таковых. Толкование представляет собой своеобразный текст на некотором языке, построенный по своим формальным законам и описывающий значение данного слова. Законы, на основе которых строятся толкования, достаточно сложны, и мы не имеем возможности здесь на них подробно останавливаться. (Читателей, которых интересуют принципы построения правильных толкований, мы можем отослать к книге Ю.Д.Апресяна «Лексическая семантика» (М., 1974) или к «Энциклопедическому словарю юного филолога» (М., 1984) — к тем статьям, которые посвящены лексикографии.) Заканчивается описание значений особой зоной иллюстраций. В ней приводятся примеры словосочетаний с данным словом в данном значении. После того как описаны все значения данного слова, составители словарей прибегают к специальному значку («ромбику»), предваряющему зону фразеологии. В ней приводятся фразеологические обороты с данным словом и даются их толкования. Впрочем, когда фразеологические обороты тесно связаны с каким-нибудь одним значением слова, они помещаются, наряду с устойчивыми сочетаниями, прямо в описание соответствующего значения.

Перечень приведенных выше зон, вообще говоря, не исчерпывает всех возможностей: есть толковые словари с зоной этимологии, бывают зоны образных иллюстраций (рисунков, таблиц) и др.

Итак, мы закончили рассмотрение требований, предъявляемых к правильно построенным словарям. Реально существующие словари могут отклоняться от идеала. В частности, иногда нарушается требование о фиксированном порядке следования зон, заполнение зон также бывает не единообразно. Важно, что каждый реальный словарь теперь, при наличии явно сформулированных точных требований, может быть оценен с точки зрения соответствия идеалу.

### **3. ПРАКТИЧЕСКИЕ ТРЕБОВАНИЯ К СЛОВАРЮ КАК ОБЪЕКТУ ИСПОЛЬЗОВАНИЯ**

Одно дело — требования к внутреннему устройству словаря, и совсем другое — практические требования к словарю как к объекту использования. Здесь не идет речь об уже упоминавшихся внешних требованиях (компактность, красивое оформление, качественное полиграфическое исполнение). Мы имеем в виду требования, выполнение которых обеспечивает обнаружение в кратчайшее время нужной пользователю словарной информации.

1. Предполагается, что словари содержат истинную информацию, т. е. не содержат ошибок.

2. Следующее требование в науке принято называть *релевантностью*. Словарь считается релевантным для данного лица, если содержит ту и только ту информацию, которая ему необходима. Если вы хотите найти в толковом словаре слово *раритет* и выяснить его значение, а в словаре этого слова не оказывается, то словарь, возможно, является хорошим, но для вас нерелевантным. Не будет для вас релевантным и такой словарь, который содержит ненужную вам избыточную информацию. Так, если вы узнаете по словарю, что «Раритет» — это название одного из книжных магазинов в Москве, эта информация для вас, ищущего значение данного слова, окажется избыточной.

Важным практическим следствием из требования релевантности является обязательность *указания адреса* словаря — то есть тех, кому он предназначен. Так, среди толковых словарей русского языка одни обращены непосредственно к носителям русского языка (как родного), а другие — к иностранцам, изучающим русский язык (как иностранный). В словарях первого типа избыточными являются сведения, касающиеся общепринятых форм поведения, специфичных для русских. Нет, скажем, надобности при слове *обед* указывать число и принятый порядок следования блюд за столом: на *первое* подают какой-либо суп (ему может предшествовать закуска), на *второе* идет горячее блюдо нежидкой консистенции, на *третье* — сладкое блюдо (десерт). В словаре, предназначенном для иностранцев, такая информация необходима; без знания ее иностранец не поймет смысла таких фраз, как *Первого не надо, мне хватит одного второго*.

Выделяют также специальные словари, предназначенные для школьников, научных работников, работников радио и телевидения и т. д. Каждый из них обладает своими особенностями, связанными с типом адресата.

3. Третье требование — *извлекаемость информации*. В значительной мере это свойство связано с правильной построенностью словаря, и прежде всего с его иерархической структурой, например алфавитный порядок позволяет легко отыскать нужное слово.

Однако одной правильной построенности недостаточно. Все словари можно условно разделить на активные и пассивные. *Пассивные* словари ориентированы только на восприятие чужой речи (чтение, понимание), а *активные* — на производство собственного текста (построение предложений, оформление мыслей на языке). Строение этих словарей должно быть разным.

Иностранец, читая русский текст и встретив в нем незнакомое слово *шел*, может обратиться за помощью к

пассивному словарю. Поэтому в таком словаре нужно предусмотреть словарную статью с входом *шел* и отсылочной интерпретацией, связывающей *шел* со словом *идти*.

Иное дело, если иностранец хочет построить русскую фразу и употребить в ней глагол *идти* в прошедшем времени. Здесь пассивный словарь ему не помощник: ведь словарную статью со входом *шел* он не найдет (даже если она содержится в словаре), поскольку сама форма *шел* ему неизвестна. Взяв в руки активный словарь, где при каждом глаголе дается множество его форм, он найдет нужную ему форму в словарной статье с входом *идти*.

Таким образом, свойство извлекаемости относительно, оно зависит от типа словаря.

Свойство извлекаемости обуславливает и форму подачи материала. Все сведения в словаре должны быть даны в явной и недвусмысленной форме; в словаре не должно быть непонятных значков, символов и сокращений; зоны должны быть отделены друг от друга удобной шрифтовой разметкой и т. д.

Внешнее требование компактности, к сожалению, часто вступает в конфликт с требованиями к форме подачи материала. Иногда для сокращения объема словаря авторы применяют особые символы, требующие специальной расшифровки. При этом «шифр» может быть достаточно сложным.

4. Последнее требование можно было бы назвать требованием рациональной избыточности. Его содержание достаточно просто. В словарь в качестве обязательных компонентов должны входить заглавие, инструкция «Как пользоваться словарем» и необходимый справочный материал: таблицы, рисунки, правила, а также некоторые энциклопедические сведения.

Может показаться, что требование рациональной избыточности вступает в противоречие с требованием релевантности. Однако это «противоречие» разрешается тем, что, во-первых, вся дополнительная информация формально не считается относящейся к основному тексту словаря, а во-вторых, в отличие от нерелевантной информации, она полезна: к ней обращаются как к вспомогательному средству для получения релевантной информации. Вспомним последний раз наш пример со словом *паритет*. Данные о том, что это слово является названием магазина, едва ли можно использовать для определения значения этого слова. В то же время рисунок, изображающий спаниеля, позволяет отличить спаниеля, например, от эрдельтерьера и тем самым точнее зафиксировать значение слова *спаниель*.



Итак, мы рассмотрели четыре важных практических требования к словарю как к полезному орудию владения языком. Это истинность, релевантность, извлекаемость информации и рациональная избыточность. Как и в ситуации с правильной построенностью, не все эти требования в равной мере важны и необходимы. Но составители словарей, явно и неявно имея в виду эти свойства, стремятся в своей деятельности к их реализации.



20. Построить классификацию имеющихся у вас дома словарей.

## II. ЛЕКСИЧЕСКИЕ ФУНКЦИИ

### 1. ЧИСЛОВЫЕ И ЛЕКСИЧЕСКИЕ ФУНКЦИИ

#### 1.А. Функция как математическое понятие

Понятие функции, как вы прекрасно знаете, входит в круг основных математических понятий (наряду с понятиями «натуральное число», «множество», «алгоритм» и др.). Однако в школе вы изучали главным образом числовые функции, т. е. определенного вида зависимости между множеством чисел. Между тем понятие функции имеет значительно более широкое применение. В частности, в науке о языке тоже говорят о функциях в понимании, близком к математическому.

#### 1.В. Функция в лексике

Ниже мы расскажем вам об одном классе таких функций, а именно о лексических функциях. Впервые их ввели в обиход крупнейшие ученые-лингвисты, специалисты в области формальной лингвистики А.К.Жолковский и И.А.Мельчук.

Начнем с примера. Рассмотрим словосочетание *категорический протест*. Что оно означает? В первом приближении оно означает 'высокую степень протеста', т. е. слово *категорический* является носителем смысла 'высокая степень' при слове *протест*. Оказывается, этот смысл можно применить (как функцию) к очень большому числу слов, причем он может выражаться самыми разнообразными способами; иными словами, существует много разных словесных выражений, передающих этот смысл. Выбор конкретного способа выражения не произволен, а зависит от «аргумента», т. е. слова или словосочетания, к которому этот смысл («функция») применяется. Например, этот смысл можно применить к слову *брюнет* ('брюнет в высокой степени'); «значением» функции, однако, здесь уже будет не *категорический брюнет* (так по-русски не говорят), а *жгучий брюнет*; смысл 'сон в



высокой степени' передается сочетанием *крепкий сон* или *глубокий сон*; смысл 'в высокой степени' от слова *негодяй* — *отъявленный негодяй*. Не следует думать, что смысл 'в высокой степени' применяется только к существительным. Например, 'болен в высокой степени' — *тяжело болен*; 'знать в высокой степени' — *знать назубок* или *знать как свои пять пальцев*.

Возьмем еще один смысл, который так же, как и смысл 'в высокой степени', выражается при очень многих словах. Это смысл 'сделать так, чтобы (был)'. Например, 'сделать так, чтобы (был)' от слова *сквер* — *разбить сквер*; 'сделать так, чтобы (был)' от слова *костюм* — *сшить костюм*; тот же смысл применительно к слову *боль* выражается сочетанием *причинить боль*.

Такие смыслы, как 'в высокой степени' и 'сделать так, чтобы (был)', стали называть лексическими функциями. Слова, к которым их применяют, называют, как и в математике, аргументами, а получившиеся словосочетания — значениями функции.

### 1.С. Отличие лексической функции от числовой

Существенно, что в отличие от числовых функций, конкретными значениями лексической функции могут быть не одно словосочетание, а сразу несколько (т.е. более чем одноэлементное множество словосочетаний). Например, значениями функций «в высокой степени» от аргумента *дождь* являются словосочетания *сильный дождь* и *проливной дождь*.

## 2. ОСОБЕННОСТИ ЛЕКСИЧЕСКИХ ФУНКЦИЙ

### 2.А. Обозначение

Лексические функции принято обозначать латинскими буквами (возможно, с индексами). Например, функцию, которую мы ранее называли 'в высокой степени', записывают латинскими буквами Magn (от латинского magnus 'большой'): Magn (брюнет) = *жгучий брюнет*; Magn (знать) = {хорошо знать; знать назубок}. Функция 'сделать так, чтобы (был)' обозначается как Caus (от латинского causa 'причина'). Таким образом, Caus (сквер) = *разбить сквер*; Caus (боль) = *причинить боль*.

### 2.В. «Склеенные» функции

Вы, наверно, уже обратили внимание на то, что значением лексической функции чаще всего бывает словосочетание или множество словосочетаний, включающих в себя слово-аргумент. Однако так бывает не всегда. Существуют так называемые «склеенные» значения функции, которые отличаются тем, что они представляют собой слова или словосочетания, не содержащие аргумента. Например, Caus (спать) = {усыпить; убаюкать}; Caus (согласиться) = {убедить; уговорить}.

Иногда значение одной и той же функции включает как «склеенное», так и «несклеенное» выражение: Magn (дождь) = {сильный дождь; проливной дождь; ливень}.

### 2.С. Типы лексических функций

**Стандартные и нестандартные лексические функции.** Может возникнуть вопрос: какие смыслы принято считать лексическими функциями и почему? По замыслу создателей теории лексических функций, к последним относятся только те смыслы, которые имеют достаточно широкую область определения и получают разнообразные лексические выражения. Например, смысл 'желтый' не является лексической функцией, поскольку этот смысл всегда можно передать присоединением к определенному слову прилагательного *желтый*. В теории лексических функций в действительности различаются так называемые стандартные и нестандартные лексические функции, при этом смыслы типа 'желтый' не относятся к стандартным лексическим функциям. В целях упрощения мы говорим здесь о лексических функциях, имея в виду лишь стандартные.

Иное дело смыслы Magn, Caus, а также Perm 'разрешить' (от латинского *permittere* 'позволить'), Inser 'начать' (от латинского *incipere* 'начать'), Cap 'начальник' (от латинского *caput* 'голова', 'глава') и многие другие. Приведем примеры

выражения этих функций: Perm (въезд) = разрешить въезд; Perm (пройти) = {позволить пройти; дать пройти; пропустить}; Inser (спать) = заснуть; Inser (дружить) = подружиться; Сар (детский сад) = заведующий(-ая) детским садом; Сар (школа) = директор школы; Сар (университет) = ректор университета; Сар (семья) = глава семьи.

**Сложные лексические функции.** Над лексическими функциями совершаются различные операции, в частности такая операция, как последовательное применение (композиция) лексических функций, которое дает в результате сложную лексическую функцию. Рассмотрим, например, функцию Figur, означающую 'фигуральное обозначение' (принятую в языке метафору) для аргумента: Figur (сатира) = жало сатиры; Figur (сон) = {объятия сна, объятия Морфея}; Figur (заговор) = нити заговора. Figur часто дает сложную функцию в композиции с Magn: так, Figur Magn (страсть) = пламя страстей; Figur Magn (отчаяние) = бездна отчаяния.

**Обратные лексические функции.** Для некоторых лексических функций определены обратные (в точном математическом смысле этого термина). Например,  $\text{Inser}^{-1}$  (заснуть) = спать;  $\text{Inser}^{-1}$  (встать) = стоять;  $\text{Inser}^{-1}$  (получить) = иметь. Эта функция, обратная функции Inser, носит название Result и означает 'результатирующее состояние', т.е. состояние, возникшее в результате данного действия — действия, обозначенного аргументом.

### 3. ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ЛЕКСИЧЕСКИХ ФУНКЦИЙ

На этом мы закончим обзор очень важного лингвистического понятия — лексической функции. Применение лексических функций в процессе автоматической обработки текста, в машинном переводе, при реферировании текста дает возможность построить компактную и обозримую систему правил, позволяет установить регулярное соотношение между исходным и преобразованным текстом. Особенно важным оказалось использование лексических функций при составлении одно- и двуязычных словарей. Аппарат лексических функций дает возможность фиксировать ранее не замеченные соотношения между словами и выявлять типичные контексты употребления лексических единиц.

Наконец, при помощи лексических функций удобно описывать фразеологические сочетания, в частности выделяя как отдельный класс те из них, которые являются значениями лексических функций.



21. Найти Magn (ветер); Magn (тьма); Magn (спать); Magn (кричать); Caus (слезы); Caus (свитер).

22. Дано: Son (корова) = мычать; Son (овца) = блять; Son

(телефон) = звонить. Найти Сон (курица); Сон (собака); Сон (колокол).  
Найти хотя бы один  $x$  такой, что: а) Сон ( $x$ ) = мяукать; то же задание для (б): б) Сон ( $x$ ) = шипеть.

23. Решите уравнения (на множестве русских слов):

а)  $\text{Magn}(x) = \text{до зубов}$ ;

б)  $\text{Magn}(y) = \{\text{иссиня-; как смоль}\}$ .

24. Вставьте пропущенные слова, установив лексическую закономерность:

а) огонь погас; война окончилась; заря догорела; ветер улегся; мода ...; интерес ...; ... отпустила;

б) отменить норму; распустить совещание; ... иллюзии; прервать сон; ... дефекты; отвести ... .

25. Наборщик случайно неверно расставил прилагательные при существительных. Помогите ему исправить ошибку и расставьте их правильно (в нужных случаях меняя род прилагательных):

<i>грубый</i>	<i>прием (гостей)</i>
<i>упорное</i>	<i>уважение</i>
<i>глубокое</i>	<i>сопротивление</i>
<i>богатое</i>	<i>наказание</i>
<i>коренной</i>	<i>опыт</i>
<i>яровая</i>	<i>болезнь</i>
<i>тяжелая</i>	<i>зависть</i>
<i>суровое</i>	<i>изменение</i>
<i>острый</i>	<i>гнев</i>
<i>горячая</i>	<i>ошибка</i>

### III. ЛИНГВИСТИЧЕСКИЕ МОДЕЛИ И СЛОВАРИ НОВОГО ТИПА

#### 1. МОДЕЛЬ И МОДЕЛИРОВАНИЕ

Выше мы привели много примеров лексических функций. Их описание и исследование составляют особый раздел науки о языке — теорию лексических функций. Использование лексических функций возможно не только для хорошей теории языка, но и для решения разного рода прикладных задач, связанных с автоматической обработкой текста.

##### 1.А. Понятие модели

Обычно, имея перед собой особую прикладную задачу (скажем, создание системы автоматического перевода, например перевода с одного языка на другой при помощи ЭВМ), исследователь берет готовую действующую модель языка, а если такой модели у него нет, он сам ее строит. Что же такое действующая модель языка?

Если бы такой вопрос вам задали о модели самолета или корабля, вы бы легко на него ответили. Модель самолета —

физический объект, который похож на самолет своим строением и внешним видом. Если бы эта модель была действующая, то ее поведение на земле и в воздухе также напоминало бы поведение самолета.

В школе все мы изучали структуру молекулы на ее модели. Глобус — это тоже модель, модель земного шара. На уроке физики вы знакомитесь со структурой атомного ядра.

Хорошая модель объекта сохраняет все свойства этого объекта, которые нужны, и не сохраняет прочих свойств. Поэтому она устроена проще, чем сам объект.

Так, модель самолета, напоминающая самолет внешне, не содержит всех тех деталей, какие есть в настоящем самолете, а если эта модель летающая, то она, по всей вероятности, не может совершать всех тех движений в воздухе, на которые способен самолет. В этой модели нет пилота и пассажиров.

Школьный глобус отражает только те свойства реальной поверхности Земли, которые изучаются на уроках географии. В частности, по этой причине он имеет форму шара, а не геоида, как Земля.

## **1.В. Модели решения задач**

Понятие модели можно распространить и на объекты другой природы. Создание моделей, или *м о д е л и р о в а н и е*, стало основным методом прикладных научных исследований. В частности, в математике создают модели решения задач. Так, уравнение, описывающее данную задачу, можно по праву считать ее моделью.

Рассмотрим две простенькие задачки.

1) Один аквариум в два раза больше по объему, чем другой. Вместе они вмещают 60 литров воды. Сколько литров воды вмещает каждый?

2) Турист прошел в первый день расстояние в два раза больше, чем во второй. А всего за два дня он прошел расстояние в 60 км. Сколько километров он прошел за каждый из двух дней?

Услышав эти задачи, вы, наверно, скажете: «Да это же одно и то же!» Но ведь в задачах речь идет о совсем разных объектах и ситуациях. Формальным подтверждением интуитивно ощущаемого тождества этих задач является то, что для них можно легко построить одну и ту же модель. Такой моделью служит уравнение  $2x + x = 60$  или, быть может, какое-нибудь равносильное ему. Это уравнение и есть модель задачи, потому что, решив его, мы решим задачу. При построении данной модели мы, во-первых, отбросили все ненужные подробности, «очистив» задачу,

а во-вторых, осуществили необходимые подстановки и замены: например, вместо слов *вместе* и *всего* мы написали знак сложения (+).

### 1.С. Модели языка

Моделируя язык или конкретное языковое явление, лингвист, занимающийся точными методами анализа естественного языка, также строит некоторую искусственную систему, «очищенную от примесей», которая во многих отношениях проще, чем язык или языковое явление, но сохраняет его важнейшие свойства. Действующая модель языка «действует» как язык: совершает преобразование одних языковых элементов в другие по определенным правилам, вводит новые элементы, выбрасывает ненужные и т. д.

Так, словообразовательная модель строит по заданному слову его производное в соответствии с правилами данной модели. Например, говорят, что слова *стол* и *дом* построены по одной модели — присоединением к производящей основе уменьшительного суффикса *-ик* (ср. аналогию с одной моделью для двух разных задач).

### 1.Д. Классы лингвистических моделей

**Анализирующие и синтезирующие модели.** Основным типом лингвистических моделей в настоящее время признаются модели, имитирующие речевую деятельность человека. Принято различать два класса таких моделей: анализирующие и синтезирующие. Конечная цель анализирующих моделей состоит в том, чтобы по данному звуковому потоку или письменному тексту определить его смысловое содержание. Синтезирующие модели делают в точности противоположное: для заданного смысла они находят способы его выражения.

Таким образом, оба класса моделей моделируют реальные интеллектуальные способности человека: а) понимание услышанного или прочитанного; б) построение отдельных высказываний и целых текстов.

**Частные модели.** Существуют, кроме того, частные анализирующие и синтезирующие модели и соответствующие им алгоритмы. Частными они называются потому, что моделируют не язык в целом, а конкретное языковое явление. Так, модель, позволяющая для каждого слова построить любую его грамматическую форму, называется моделью синтеза словоформы. Напротив, модель анализа словоформы дает возможность для каждой реально встречающейся в тексте словоформы установить ее начальную (основную) форму и определить грамматическое значение (т. е. как бы произвести полный морфологический разбор словоформы).

## 2. СЛОВАРИ НОВОГО ТИПА

### 2.А. Словарь как центральная часть лингвистических моделей

Центральной частью синтезирующих и анализирующих моделей языка является словарь. Поскольку модель, вообще говоря, предполагает в дальнейшем алгоритмическую реализацию, это обстоятельство накладывает на словарь определенные требования. К числу важнейших требований относятся: 1) этот словарь должен быть правильно построенным (в том смысле, в каком об этом говорилось выше); 2) должна быть соблюдена максимальная точность подачи сведений, причем все сведения должны быть представлены в словаре в явном виде; 3) в словаре должна содержаться та и только та информация, которая используется в правилах анализа или синтеза.

Для успешной реализации этих требований составители подобных словарей прибегают к помощи математического аппарата: математической и логической символики, аппарата лексических функций, записи информации в виде логических правил.

Таким образом, данные словари по праву можно отнести к словарям нового типа, порожденным веком информационных технологий и ЭВМ. В частности, исторически задачи, связанные с созданием промышленных систем машинного перевода с одного языка на другой, вызвали необходимость в построении нового, «толково-комбинаторного» словаря русского языка. Его создатели поставили своей целью дать максимальное количество сведений (в идеале все) о каждом слове, включенном в словарь.

### 2.В. Устройство статьи толково-комбинаторного словаря

Посмотрим более подробно, как устроена какая-нибудь словарная статья толково-комбинаторного словаря.

Возьмем, например, слово *ветер*. Словарь подчеркивает особенности образования и произношения форм этого слова (предложный падеж - *о вѣтрѣ*, но *на ветру́*, множественное число *вѣтры* и разговорное *ветра́* и др.). Основной зоной словарной статьи является толкование. Особенность толкования состоит в том, что оно всегда строится в виде правила, в левой части которого помещается толкуемое слово вместе с минимальным контекстом его употребления (причем полноточные слова заменены переменными), а в правой части стоит текст, описывающий бытовое представление о значении слова. *Ветер с Y-а в сторону Z-а*  $\approx$  'достаточно сильное перемещение воздуха, в направлении от Y-а к Z-а, происходящее вблизи земной поверхности, захватывающее достаточно широкое пространство'. Значение слова *ветер* иллюстриру-



ется примерами: *Сегодня на улице страшный ветер; Первые порывы ветра по палубам кораблей; Ветер по морю гуляет и кораблик подгоняет* (Пушкин).

Со словом *ветер* у носителей русского языка связаны устойчивые ассоциации: свободы (ср. *вольный, как ветер*), быстроты (ср. *быстрый как ветер*), несерьезности (ср. *ветер в голове; ветренный человек; бросать слова на ветер*) и др. Все эти ассоциации указываются в отдельной зоне словарной статьи.

В особой зоне «сравни» даются слова, смежные по значению: *дождь; снег; сквозняк* и т. п.

Очень важно указать способы оформления значений переменных в толкуемом выражении. В частности, переменная *У* имеет значение 'откуда перемещается <ветер>' и оформляется одним из двух способов: 1) предложно-падежной формой со значением места (*ветер с гор, с юга*); 2) прилагательным (*южный ветер*).

Очень богата зона лексических функций и их выражений. В ней приводятся синонимы (полные и частичные: *ветр, вихрь*), антонимы (*безветрие, штиль*), обобщающее слово (*атмосферное явление*), указываются основные производные данного слова (*ветренный, ветряной, проветрить, обветриться, ветронепроницаемый*), значения лексических функций Magn: {резкий ветер; сильный ветер; порывистый ветер; шквал}, Anti Magn: {легкий ветер; слабый ветер; ветерок}, Son Magn: {выть; завывать; реветь; свистеть}, Son Anti Magn: {шелестеть}, названия типовых действий, совершаемых ветром (*разгуляться, ослабеть, стихнуть, бушевать, свирепствовать, подниматься, изменить направление, дуть, повеять* и др.), возможно, с указанием специальных условий реализации (когда ветер дует в разные стороны и не встречает препятствий на своем пути, то говорят, что он *гуляет*, а если дует в одну сторону, то *проносится*), перечисляются глаголы, подчеркивающие способность ветра быть разрушающей силой (*разогнать тучи, развеять туман, сбить с ног человека, сорвать шапку, продуть насквозь*), а также слова, указывающие на пространственные координаты ветра (абсолютные и относительные: *переменный ветер, попутный ветер, встречный ветер*), и т. п.

Отдельное место посвящено энциклопедической информации. Указываются виды ветров: а) по географическому характеру местности (*бриз* — береговой ветер, дующий днем с моря, а ночью с суши; *муссон* — ветер, имеющий направление в зависимости от времени года: зимой — с материка на океан, летом — с океана на материк, и др.); б) по местности (названия местных ветров: *свежак* — ветер у берегов Черного моря; *мистраль* — холодный ветер на юге Франции; *самум* — знойный сухой ветер в пустынях Азии и Африки и др.); в) по

направлению (*норд* — северный ветер; *зюйд* — южный ветер; *ост* — восточный ветер; *вест* — западный ветер; и производные от них: *норд-ост*, *зюйд-ост* и т. п.). Здесь же рассказывается о физических параметрах ветра и о том, как они измеряются, о приборах и приспособлениях, действие которых основано на свойствах ветра (*флюгер*, *парус*, *ветряная мельница*); приводятся сведения из мифологии, легенд и других культурных источников об обозначении ветра и связанных с ним явлений: *Эол* — бог ветра в греческой мифологии; *борей*, или *аквилон* — поэтические названия холодного северного ветра (у греков и римлян соответственно); *зефир* — легкий теплый ветер (у древних греков).

В следующей зоне приводятся языковые обозначения ветра, учитывающие сразу несколько его свойств: *вихрь*; *смерч*; *поземка*; *метель*; *вьюга*; *пурга*; *буря*; *шторм*; *буран*; *ураган*.

Заключает словарную статью зона фразеологии, в которой истолкованы наиболее употребительные фразеологизмы со словом *ветер*: *как ветром сдуло* = 'неожиданно и быстро исчез'; *каким ветром тебя занесло?* = 'говорящий удивлен неожиданным появлением собеседника'; *ищи ветра в поле* = 'искать бесполезно — того, кого (что) ты ищешь, уже нет'. Во всех этих изречениях обыгрывается связь ветра с быстротой перемещения. Ассоциация ветра с переменчивостью иллюстрируется фразеологизмами *держат нос по ветру*; *знать куда (откуда) ветер дует*; а с несерьезностью — фразеологизмами *ветер в голове*; *бросить слова на ветер* и т. д. С разрушительной силой ветра связано изречение *посеешь ветер — пожнешь бурю*.

Уже по одному этому примеру словарной статьи вам, наверно, понятно, насколько большой и разнообразный объем сведений она содержит. Может возникнуть вопрос: нет ли здесь избыточности? Ведь обычно в толковые словари помещают меньше сведений, а люди ими успешно пользуются. Ответом на этот вопрос может служить важное положение о том, что словарь всегда создается с расчетом на конкретного адресата. Когда этим адресатом является человек, составители, естественно, опираются на его интеллектуальные и творческие способности, в частности на логическую способность действовать по аналогии и умение обобщать, а также на его интуитивное владение языком. Иное дело, когда словарь рассчитан на бездушную машину. Очевидно, что здесь нельзя обойтись без всех сведений, которые сообщаются в словарях типа толково-комбинаторного. Ведь без них машина не сможет понимать и создавать русские тексты. Все сведения, представленные в словарных статьях таких словарей, непосредственно используются в алгоритмах автоматического анализа и синтеза текста.

## 2.С. Машинные словари

По мере внедрения электронно-вычислительной техники в различные сферы человеческой деятельности возникла необходимость в создании интеллектуальных систем обработки информации с использованием ЭВМ. Все такие системы в качестве двух составляющих имеют лингвистическое и математическое обеспечение. Одна из сторон лингвистического обеспечения — это создание машинных словарей, т. е. словарей на машинных носителях.

Машинные словари необходимы для обнаружения и исправления ошибок при вводе информации в ЭВМ, для поиска информации, для построения автоматических классификаций, для составления рефератов текстов. Составление машинных словарей невозможно без обработки «обычных» словарей, созданных людьми и для людей. Например, на базе грамматического словаря создается машинный словарь основ, где дается информация о типе словоизменения и словообразования. Кроме того, бывают полезны машинные словари словоформ (прямые и обратные). Существуют особые машинные процедуры, позволяющие осуществлять автоматическую обработку текста, который содержит слова, отличные от слов данного словаря. Во Всесоюзном институте научной и технической информации созданы разные виды машинных словарей для многих тематических областей: транспорт, металлургия, химия, экология и т. д.

### 2. D. Автоматизация лексикографических работ

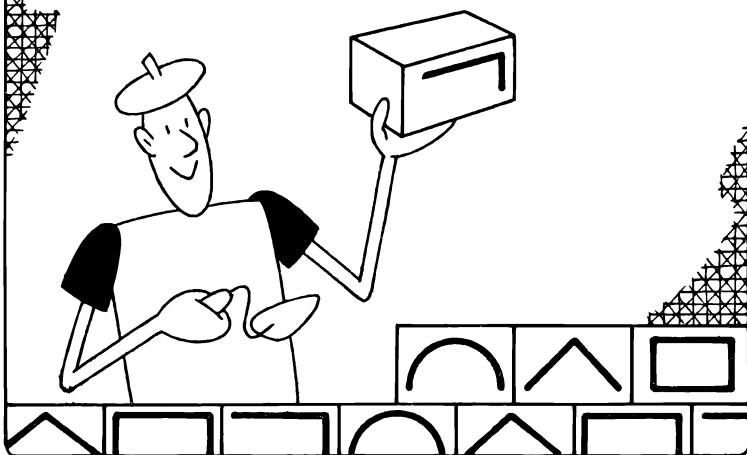
Созданием машинных словарей отнюдь не исчерпывается применение ЭВМ в лексикографии. Важной задачей является автоматизация подготовительной работы — составление карточек, словарная обработка поступающих на вход текстов, создание частотных словарей и конкордансов. (Конкорданс некоторого текста — это словарь, в котором для каждого слова из данного текста указываются все встретившиеся употребления этого слова в данном тексте вместе с ближайшим окружением.)

Очень важной областью использования ЭВМ в лексикографии является создание разного рода вспомогательных текстов к словарям. Среди них различные указатели, списки источников, списки помет, сокращений и т. д.

Итак, мы с вами познакомились с некоторыми словарями, появление которых вызвано, с одной стороны, развитием теоретической науки, а с другой стороны — широким применением ЭВМ в языкознании.

Глава четвертая

# Формальное строение слова



# 1. ФОРМАЛЬНЫЙ ЯЗЫК МОРФЕМИКИ

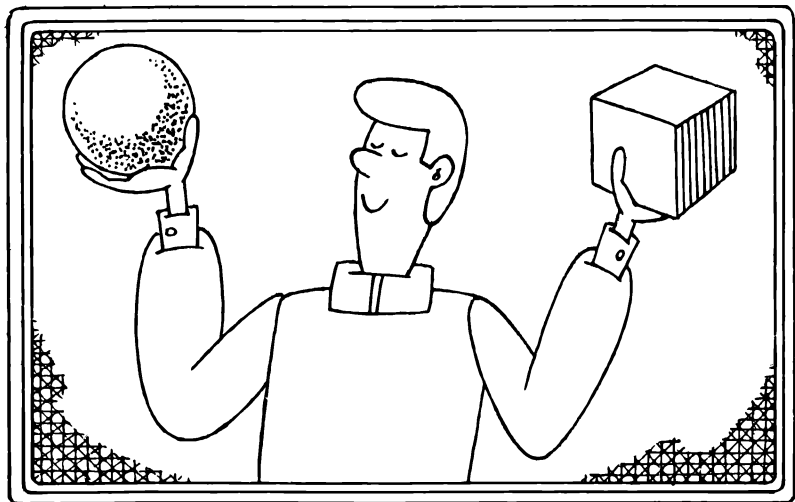
## 1. ФОРМАЛЬНЫЕ И НЕФОРМАЛЬНЫЕ ОПРЕДЕЛЕНИЯ

### 1.А. Математические определения

**В**се люди с самого раннего детства отличают кубик от шарика. Но попробуйте задать ребенку вопрос: в чем именно состоит это отличие? Думаем, что его ответ вас никак не удовлетворит. В чем же дело? Ведь все так хорошо знают эти предметы, много раз их видели, в детстве играли с ними. Здесь мы сталкиваемся с парадоксальным, на первый взгляд, явлением. Отличить кубик от шарика может всякий, а точно указать, в чем состоит отличие, может только тот, кто изучил курс геометрии и познакомился в школе с понятиями куба и шара, и прежде всего с их определениями.

Изучая школьный курс математики, мы впервые узнаем, что **определение** — это текст, описывающий существенные свойства данного предмета, отличающие его от всех других предметов (или данное понятие от всех других понятий).

Умение дать правильное определение относится к числу наиболее важных умений. Без овладения этим навыком невозможна никакая научная деятельность. Какой бы вы учебник ни открыли, вы всегда найдете там десятки определений, описывающих объекты данной дисциплины.



## 1.В. Лингвистические определения

Содержатся определения и в учебнике русского языка. Откроем, например, в учебнике «Русский язык» раздел «Состав слова» и прочтем: «Суффикс — это значимая часть слова, которая находится после корня и обычно служит для образования слов»; «Окончание — это изменяемая значимая часть слова, которая образует форму слова и служит для связи слов в словосочетании и предложении».

Чувствуется, что определения в учебнике русского языка чем-то отличаются от определений, услышанных нами на уроках математики. Это отличие состоит в том, что математические определения формальные, а определения, содержащиеся в учебниках русского языка, — нет.

### 1.С. Формальное устройство формальных определений

Формальное определение — это определение, удовлетворяющее требованиям, которые можно условно разделить на два типа: требования, относящиеся к форме текста определений, и требования, относящиеся к его содержательной стороне.

По форме определение — это текст вида  $A \rightarrow B$  (или легко сводимый к таковому). Здесь  $A$  обозначает определяемое понятие, а  $B$  — его объяснение, или, как принято говорить, толкование. Стрелка « $\rightarrow$ » — это символ связи между левой и правой частью, которая обычно выражается в русском языке словами *это, называется* или знаком тире (—).

Требования к содержанию текста определения еще более строгие. Все понятия, используемые в правой части, т. е. в толковании, должны либо быть заимствованы из исходного списка неопределяемых понятий, либо быть определены заранее, вне всякой связи с данным определением.

Например, в языке геометрии формальным определением может служить такой текст: «Смежными углами называются два угла, у которых одна сторона общая, а две другие образуют прямую линию». Понятие 'прямой линии' здесь относится к числу неопределяемых, а понятия 'угла' и 'стороны угла' считаются определенными ранее. В этом тексте встречаются также слова *один, два, который и образовать*, постоянно использующиеся в текстах определений; их обычно называют синтаксическими константами. Приведем еще примеры синтаксических констант: *всякий, существует, ровно один, если... то, можно* и др.

### 1.Д. Формальные определения в лингвистике

До недавнего времени круг задач, который решался наукой о языке, позволял обходиться неформальными

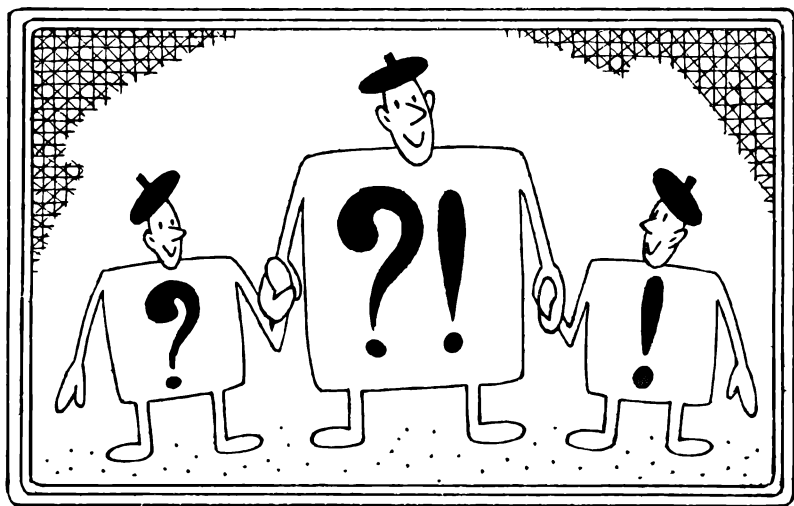
определениями, не претендующими на высокую точность и строгость. Однако развитие лингвистической науки, научно-техническая революция и прежде всего совершенствование электронно-вычислительной техники поставили перед лингвистами новые задачи: создание алгоритмов решения лингвистических проблем (автоматическая обработка и переработка текста, машинный перевод), конструирование интеллектуальных устройств (роботов), понимающих команды на естественном языке, систем обращения человека с животными. Оказалось, что их решение предполагает в качестве необходимого условия построение целой системы формальных определений языка лингвистики. Роль такой системы в языке лингвистики сходна с ролью алгоритмических языков, с которыми вы сталкиваетесь в курсе информатики и электронно-вычислительной техники.

## 2. ОСНОВАНИЯ СИСТЕМЫ ФОРМАЛЬНЫХ ОПРЕДЕЛЕНИЙ

Давайте теперь вместе попытаемся построить фрагмент такой системы для одного раздела науки о языке — морфемике, т.е. учения о значимых частях слова. На основе этого фрагмента в дальнейшем можно будет дать пример содержательного лингвистического алгоритма — алгоритма разбора слова по составу.

### 2.А. Исходные неопределяемые понятия

Как и любая формальная система, система, которую мы строим, должна быть основана на исходных (неопределяемых



в рамках данной системы) понятиях. К их числу относятся, во-первых, центральные лингвистические понятия: языкового знака, означаемого и означающего языкового знака, контекста. Кроме того, в число исходных понятий входят общелогические и математические термины, такие, как 'алгоритм', 'множество', 'элемент множества', 'операция', 'процедура', 'последовательность', 'переменная', 'отношение эквивалентности' и др. Напомним, что необходимые общелогические и математические понятия уже были так или иначе упомянуты в I главе.

Особую роль в дальнейшем описании будет играть такое важное неопределяемое общенаучное понятие, как 'состоять из'. Вспомним примеры употребления этого понятия для знаков из главы I: «Предложение *Наступила зима* состоит из двух слов: *наступила* и *зима*; Знак ?! состоит из знаков ? и !». Составные части языкового знака мы называем **з н а ч и м ы м и ч а с т я м и**.

С самого начала хотим предупредить: не следует думать, что можно построить только одну систему формальных понятий морфемики. Из бесконечного множества теоретически возможных формальных систем мы выбираем ту, которая максимально приближена к школьному курсу и практически удобна, а в качестве исходных понятий используем только те, которые уже были ранее введены в нашей книге.

## **2.В. Требования к построению системы**

При построении системы мы будем следить за тем, чтобы выполнялся ряд условий. Прежде всего будем следить за тем, чтобы не возникали «логические крути», т. е. ситуации, когда одно понятие определяется через другое, а второе, в свою очередь, через первое. Это условие является очевидным следствием упомянутого ранее требования к содержанию формальных определений. Далее, мы будем избегать синонимии. Дело в том, что использование синонимов в текстах определений мешает установить, одинаковые перед нами определения или разные: приходится каждый раз объяснять, что данные слова имеют одинаковые значения, а это делает определение очень громоздким. Третье условие — отсутствие многозначных терминов. Использование термина попеременно то в одном, то в другом значении создает нежелательную путаницу. У учителя можно спросить, какое содержание он вкладывает в термин в каждом конкретном случае, но в книге ответа на этот вопрос может и не быть.

В частности, уже по названным причинам обычный русский язык не может служить формальным языком: ведь мы знаем, сколь богат русский язык синонимами и многозначными



словами. Между тем желательно, чтобы формальный язык был максимально приближен к естественному языку. О причинах этого вы уже не раз слышали на уроках информатики, и здесь мы подробно это положение развивать не будем.

Таким образом, конструируемый язык является своеобразной разновидностью языка нашего повседневного общения. Его своеобразие заключается в том, что в нем нет синонимов и многозначных слов, а тексты на этом языке не содержат логических кругов.

### 3. ФОРМАЛЬНАЯ СИСТЕМА ОПРЕДЕЛЕНИЙ ОСНОВНЫХ ПОНЯТИЙ МОРФЕМИКИ

Описание формальной системы морфемики мы построим следующим образом. Определяемое понятие дается в р а з р ы д к у. Как правило, после текста определения идет поясняющий его пример. На рисунках указываются последовательно один за другим все основные термины, которые были нами определены к настоящему моменту, и термин, определяемый в этом месте текста. Некоторые определения мы предложим вам дать самостоятельно как упражнение.

#### 3.А. Исходные элементы системы

**Элементарный знак.** Начнем с определения понятия 'элементарный знак'.

Языковой знак называется э л е м е н т а р н ы м, если он не состоит ни из каких других знаков.

#### **0<sub>1</sub>** элементарный знак

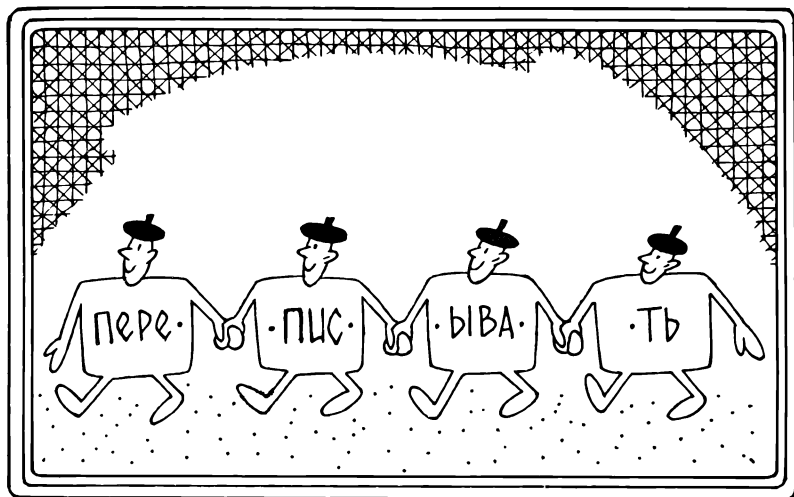
Например, слово *переписывать* не является элементарным, поскольку оно состоит из значимых частей *пере-*, *-пис-*, *-ыва-* и *-ть*, которые представляют собой элементарные знаки. Очевидно также, что знак ?! не является элементарным, а знаки ? и ! — элементарные.

**Морф.** Существует особое обозначение элементарных знаков в составе слова. Элементарная значимая часть слова называется м о р ф о м.

#### **0<sub>2</sub>** элементарный знак, морф

Морфами являются значимые части *пере-*, *-пис-*, *-ыва-* и *-ть* в слове *переписывать*. Но эти же морфы могут входить и в состав других слов, например *переделка*, *писатель*, *прочитывал*, *учить*.

Важным свойством морфа является свойство иметь одно и то же означаемое (значение) во всех случаях его употребле-



ния. Так, *пере-* в словах *переделывать*, *перечитывать*, *пересмотреть* имеет значение повторности действия ('делать заново', 'читать еще раз', 'снова рассмотреть'). Морф *пере-*, выступающий в словах *перезаниматься*, *перетрудиться* и т. п., имеет другое значение: с его помощью действие характеризуется как 'произведенное в чересчур большом количестве', т. е. перед нами два разных морфа *пере-*, имеющих одно и то же означаемое.

Бывает (и довольно часто) обратная ситуация, когда два разных морфа имеют одно означаемое, но разные означающие, т. е. являются синонимами. Например, *-иц-* в слове *волчица* и *-их-* в слове *зайчиха* оба означают 'самка'.

**Нулевой морф.** Введем теперь понятие нулевого морфа. Формальное определение этого понятия отсылает нас к понятию нулевого знака, введенному в главе I: *нулевой морф* — это нулевой языковой знак, являющийся морфом.

### **0<sub>3</sub> элементарный знак, морф, нулевой морф**

Дадим этому формальному определению содержательные пояснения. Начнем с примера. Возьмем слово *стол* и будем изменять его по падежам: *стола*, *столу* и т. д. Морф *-а* в форме *стола* указывает на то, что это форма родительного падежа единственного числа, т. е. по его наличию в слове мы узнаем, что перед нами родительный падеж. Морф *-у* указывает на дательный падеж единственного числа и т. д. Какой же

элемент в самом исходном знаке *стол* нам указывает на то, что перед нами именительный падеж единственного числа? Очевидно, отсутствие на конце этого слова морфов *-а*, *-у*, *-ом* и т. д. Таким образом, это отсутствие является здесь значимым. Во всех подобных случаях и говорят о нулевых морфах как разновидностях нулевых знаков (условное обозначение  $\emptyset$ ). Другими словами, слово *стол* состоит из двух морфов: *стол-* и  $\emptyset$ , подобно тому как слово *стола* состоит из морфов *стол-* и *-а*.

Естественно, что два разных нулевых морфа могут различаться только значением.



26. Доказать, что в русском языке нулевых морфов более одного.

Если вы решили эту задачу или прочли решение и разобрались в нем, подумайте еще над несколькими контрольными вопросами-задачами:

Доказать, что в форме *сядь* есть еще один нулевой морф. Каково его значение?

Привести примеры нулевых морфов у прилагательных, указав их значение.

Есть ли нулевые морфы у наречий? Если да, то каково их означаемое? Если нет, то почему?

В русском языке есть нулевой морф со значением 'прошедшее время'. Приведите примеры слов, в которых он выделяется (можем подсказать читателю, что таких слов немало).

### 3.В. Формальное уточнение понятия «разбор слова по составу»

С помощью понятия морфа можно точнее сказать, что значит «разбор слова по составу». **Р а з о б р а т ь с л о в о п о с о с т а в у** — значит установить, из каких морфов оно состоит и каков порядок их следования в слове.

Казалось бы, мы уже имеем все, что нужно для построения алгоритма разбора слова по составу. И действительно, в простых случаях можно было бы ограничиться только что введенными понятиями, а разбор слова по составу производить примерно так.

Возьмем слово *переписывать*. Произведя поиск по нужному словарю (который может быть записан в памяти ЭВМ или в памяти человека), находим слово *переписывал*. Эти слова различаются и значением (одно — неопределенная форма, другое — прошедшее время), и внешней стороной: в первом случае имеется последовательность *-ть*, которая во втором заменена на *-л*. Естественно считать, что *-ть* как раз и обозначает неопределенную форму, а *-л* — прошедшее время.

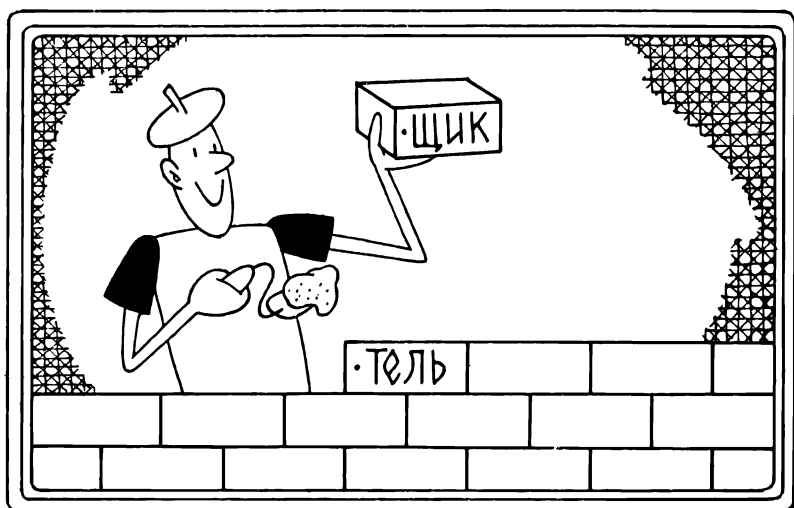
Это дает возможность выделить *-ть* и *-л* как отдельные морфы. Сравнивая *переписывать* и *переписать*, выделяем морф *-ыва-* со значением несовершенного вида и морф *-а-* со значением совершенного вида. Аналогичным образом выделяем морф *пере-* (ср. *писать*) и морф *-пис-* (ср. *переписывать*).

#### 4. MORFOЛОГИЧЕСКИЕ ОТНОШЕНИЯ И ОПЕРАЦИИ

##### 4.А. Чередование как отношение и как операция

Такой простой алгоритм не срабатывал бы, однако, в очень многих случаях. Возьмем, например, слово *бегу*. Сравнив это слово со словом *бежит* и выделив общую цепочку *бе-*, наш воображаемый алгоритм дал бы, очевидно, неприемлемый результат, отнеся цепочку *-гу* и *-жит* к морфам (со значением, соответственно, 1-го лица единственного числа и 3-го лица единственного числа). Стремление избежать таких случаев приводит к необходимости ввести важное понятие чередования.

Сначала введем вспомогательное понятие: *с т а н д а р т н ы й х—х*, встречающийся во многих случаях. Например, к стандартным морфам, обозначающим профессии в русском языке, относятся морфы *-щик* и *-тель* (ср. *регулировщик*, *кровельщик*, *наборщик*, *краснодеревщик*, *каменщик*, *учитель*, *золотоискатель*, *смотритель*, *воспитатель*), а к нестандартным — *-ач* (как в слове *ткач*); стандартным написанием является *шу* (ср. *шуба*, *шуметь*, *шутить*), а нестандартным — *ию* (как в слове *парашю*).



О п е р а ц и я ч е р е д о в а н и я — это такая стандартная подстановка в морфе, которая не меняет его означаемого. Таким образом, в результате применения этой операции получаем морф с другим означаемым, но имеющий то же означаемое, что и исходный.

**0<sub>4</sub>    элементарный знак, морф, нулевой морф,  
чередование (операция)**

О двух морфах, один из которых может быть получен из другого в результате применения операции чередования или последовательности таких операций, будем говорить, что они находятся в о т н о ш е н и и ч е р е д о в а н и я. Для простоты (и это тоже входит в определение) будем также считать, что любой морф находится в отношении чередования сам с собой.

**0<sub>5</sub>    элементарный знак, морф, нулевой морф,  
чередование (операция), чередование (отношение)**

Приведем примеры чередований. Слово *лицо* состоит из двух морфов: *лиц-* (со значением 'передняя часть головы человека') и *-о* (со значением 'именительный падеж единственного числа'); слово *личико* — из трех морфов: *лич-* (с тем же значением, что и у *лиц-*), *-ик-* (со значением уменьшительности или ласкательности) и *-о* (тот же морф, что и в слове *лицо*). Поскольку морфы *лиц-* и *лич-* имеют одно и то же значение и подстановка *ц~ч* стандартна в русском языке (ср. *отец* — *отеч(еский)*, *улиц(а)* — *улич(ный)*, *границ(а)* — *гранич(ить)*, *лестниц(а)* — *лестнич(ный)*), *лиц-* и *лич-* находятся в отношении чередования, а подстановка *ц~ч* является операцией чередования.

Также находятся в отношении чередования морфы *сон-* (в слове *сон*) и *сн-* (в слове *сна*): здесь *о* чередуется с нулем; эта подстановка стандартна (ср. *шов* — *шва*, *ров* — *рва* и т. д.). Напомним, что гласный *о*, чередующийся с нулем, называется б е г л ы м г л а с н ы м.

Двухэлементную последовательность операций чередования можно проиллюстрировать на примере морфов *ид-* (в слове *иду*) и *-й-* (в слове *пройти*): 1) стандартная подстановка фонетического нуля на место *д* (ср. *вед(у)* — *ве(ла)*, *бред(у)* — *бре(ла)*); 2) стандартная подстановка *й* на место *и* (ср. *поймка* — *поймать*; *ни(л)* — *нь[й](у)*). Тем самым *ид-* и *-й-* находятся в отношении чередования.

Трехэлементная последовательность представлена морфами *-сыл-* (в слове *посылка*) и *ш[л']-* (в слове *шлю*).



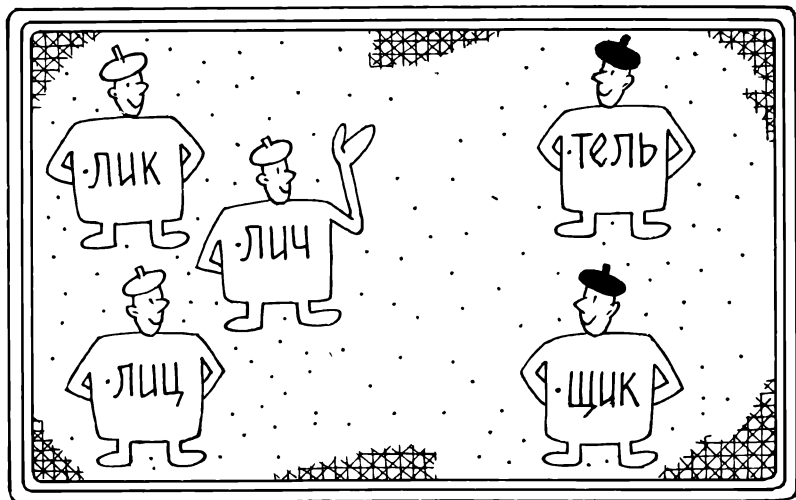
27. Выписать эту последовательность чередований в явном виде.

Не находятся в отношении чередования морфы *-тель* и *-щик* (со значением производителя действия), поскольку в русском языке нельзя указать последовательность стандартных подстановок, посредством которых достигался бы переход от одного из этих морфов к другому.

28. Доказать, что отношение чередования является отношением эквивалентности (т.е., напомним, является рефлексивным, симметричным и транзитивным).

В силу задачи 28, т.е. в силу того, что отношение чередования является отношением эквивалентности, к нему можно применить теорему о разбиении из главы 1 — о связи отношения эквивалентности с разбиением множества на классы. Получаем, что множество морфов разбивается на классы эквивалентностей по отношению чередования, так что в один класс попадают те и только те морфы, которые связаны отношением чередования. Например, в один класс попадут морфы *лик-*, *лиц-*, *лич-*. В разные классы попадут морфы *-тель* и *-щик*, а также *пере-* со значением 'заново' и *пере-* со значением 'чересчур'.

Вы, конечно, понимаете, что полученные классы эквивалентности удобно как-то назвать. На практике обычно не выбирают новое имя, а используют означающее одного из морфов данного класса. Чтобы отличить имя класса морфов от имени одного морфа — элемента класса, имя класса записывают заглавными буквами. Например, класс {*лик-*; *лиц-*; *лич-*} можно обозначить *ЛИЦ*.



Теперь можно ввести едва ли не самое важное понятие морфемики — понятие морфемы. Строгое определение этого понятия выглядит довольно сложно, а именно: *морфема* — это языковой знак, такой, что его означающее — имя класса эквивалентных (по отношению чередования) морфов, а означаемым является означаемое любого из этих морфов.

**0<sub>6</sub>** **элементарный знак, морф, нулевой морф, чередование (операция), чередование (отношение), морфема**

Поясним понятие морфемы примерами. Класс {лик-; лиц-; лич-} = *ЛИЦ* задает морфему <*ЛИЦ*, 'лиц'> (условное обозначение — *ЛИЦ*-); класс {пере<sub>1</sub>-} = *ПЕРЕ*<sub>1</sub>- — морфему *ПЕРЕ*<sub>1</sub>- = <*ПЕРЕ*<sub>1</sub>-, 'заново'>.

Условимся говорить, что морфы {лик-; лиц-; лич-} представляют морфему *ЛИЦ*- (или являются ее представителями). Аналогично будем называть морф *пере*<sub>1</sub>- представителем морфемы *ПЕРЕ*<sub>1</sub>- и т.п.

Любопытно поставить вопрос: сколько представителей может быть у одной морфемы? Ясно, что не менее одного. Приведем примеры морфем, представленных одним морфом: у в словах *уползти*, *унести*; двумя — *рук*-/*руч*- (морфема *РУК*-) в словах *рука*, *ручной*; тремя — *кус*-/*кус'*-/*куш*- (морфема *КУС*-) в словах *кусать*, *укусить*, *укушу*.



29. Доказать, что существуют морфемы более чем с тремя представителями.

30. Доказать, что морфы, представляющие разные морфемы, не могут быть связаны отношением чередования.

#### 4.В. Виды морфем и отношения между ними

Поскольку мы говорим о морфах как о представителях морфем, можно считать, что слово состоит из морфем. Например, считать, что слово *личико* состоит из трех морфем: *ЛИЦ*-, *-ИК*- и *-О*.

Встает естественный вопрос: какие виды морфем существуют в языке? Сразу дадим на него ответ: корни, суффиксы, приставки и окончания.

Дать формальное определение понятию *корня* весьма сложно. Для простоты мы будем считать понятие корня неопределяемым. Укажем лишь наиболее важное его свойство: в составе каждого слова можно выделить по меньшей мере одну морфему, являющуюся корнем. Говоря совсем уже неформально, каждое слово имеет хотя бы один корень. Помимо корня-морфемы часто рассматривают и корни-

морфы, причем соотношение между ними такое же, как между любой морфемой и представляющими ее морфами.

Исходя из понятия корня, определим понятие корневого гнезда. **Корневое гнездо** — это множество слов, имеющих общий корень. Слова, входящие в одно корневое гнездо, называются **родственными** (по корню). Приведем примеры родственных слов: слово *пар* родственно словам *паровой*, *парник*, *испарина*, *выпаривание*, *испаряться*, *запарка*, *парилка* и др. Все эти слова входят в одно корневое гнездо — {*пар*; *паровой*; *парник*; *испарина*; *выпаривание*; *испаряться*; *запарка*; *парилка* ..}.

**0<sub>7</sub>**, **0<sub>8</sub>** **элементарный знак, морф, нулевой морф, чередование (операция), чередование (отношение), морфема, корневое гнездо, родственные слова (по корню)**

Из данного определения вытекает следствие: любые два родственных слова имеют по меньшей мере одну общую морфему (но не обязательно общий морф!). Так, в приведенном гнезде все слова имеют общую морфему *ПАР-*, но слова *пар* и *парилка*, хотя и являются родственными, не имеют общих морфов (в слове *пар* корнем является морф *пар-*, а в слове *парилка* — *пар'-*).

А в какое гнездо входит слово *пароход*? Очевидно, что таких гнезд два. Одно гнездо мы уже рассмотрели, другое гнездо включает слова *ход*, *ходок*, *ходики*, *ходули*, *восхождение*, *входить*, *поход*, *доходчивый*, *перехаживать* и др. На примере этого слова мы видим, что существуют слова, входящие более чем в одно корневое гнездо. Это слова, имеющие в своем составе несколько корней. Такие слова называются **сложными**.

**0<sub>9</sub>** **элементарный знак, морф, нулевой морф, чередование (операция), чередование (отношение), морфема, корневое гнездо, родственные слова (по корню), сложные слова**

**Окончанием** принято называть морфему (или морф) в составе слова, которая служит для связи элементов в предложении. Окончание, в отличие от корня, есть не у всех слов. Например, в составе наречий типа *вчера*, *назад*, междометий (*эх*, *ай*), частиц (*даже*, *только*) и др. окончание не выделяется, тогда как в существительных *вода*, *окно*, а также *дом* окончание есть, причем в слове *дом* оно представлено нулевым морфом.



Если в составе слова есть окончание, то часть этого слова без окончания называется **о с н о в о й**. Если в составе слова окончания нет, то его основа по определению совпадает с самим этим словом.

**0<sub>10</sub>** , **0<sub>11</sub>** **элементарный знак, морф, нулевой морф, чередование (операция), чередование (отношение), морфема, корневое гнездо, родственные слова (по корню), сложные слова, окончание, основа**

Основа может состоять как из одной, так и из нескольких морфем.

Примеры:

Слово	Наличие окончания	Основа	Морфемный состав слова	Морфемный состав основы
вперед	нет	вперед	В+ПЕРЕД	В+ПЕРЕД
вчера	нет	вчера	ВЧЕРА	ВЧЕРА
сдуру	нет	сдуру	С + ДУР + У	С + ДУР + У
стеной	есть	стен	СТЕН+ОЙ	СТЕН
дом	есть	дом	ДОМ + Ø	ДОМ
какой-либо	есть	как...-либо	КАК+ОЙ+ЛИБО	КАК+...+ЛИБО
пятьдесят	есть	пят'...десят	ПЯТ'+ Ø + ДЕСЯТ + Ø	ПЯТ'+...+ДЕСЯТ
двадцать	есть	двадцать	ДВ+А+ДЦАТ'+ Ø	ДВ+А+ДЦАТ'
пароход	есть	пароход	ПАР+О+ХОД+ Ø	ПАР+О+ХОД
петь	есть	пе	ПЕ+ТЬ	ПЕ
судиться	есть	суди...ся	СУД'+И+ТЬ+СЯ	СУД'+И+...+СЯ
слепого	есть	слеп	СЛЕП+ОГО	СЛЕП

В русском языке не существует морфемы, которая одновременно была бы и корнем, и окончанием. Отсюда следует, что корень всегда входит в состав основы. Первое утверждение, несмотря на кажущуюся очевидность, доказать очень трудно. Для доказательства следовало бы иметь список всех корней русского языка с указанием типов их означаемых.

По аналогии с понятием родственных (по корню) слов легко определить понятие родственных слов (по основе), или родственных основ. Оно будет в дальнейшем использовано в приложении к данной главе. Там же используется понятие 'корневого гнезда основ' (попробуйте построить определения этих понятий сами).

Понятия 'суффикса' и 'приставки' определяются исходя из понятий 'корня' и 'основы слова'. Для простоты изложения мы будем строить определения для слов, содержащих только один корень. Для сложных слов формальные определения суффикса и приставки громоздки, и мы их приводить не будем.

**Суффикс** — это морфема (или морф) в составе основы, расположенная после корня. **Приставка**, или **префикс**, — это морфема (или морф) в составе основы, расположенная перед корнем.

**0<sub>12</sub>**, **0<sub>13</sub>** **элементарный знак, морф, нулевой морф, чередование (операция), чередование (отношение), морфема, корневое гнездо, родственные слова (по корню), сложные слова, окончание, основа, суффикс, приставка (префикс)**

В составе слова может быть больше одного суффикса или приставки. Например, в глаголе *выпроводить* две приставки: *вы-* и *про-*; в слове *учительница* — три суффикса: *-и-*, *-тель-*, *-ниц-*.

С другой стороны, приставки и суффиксы есть не в каждом слове: в слове *подпол* нет суффикса, в слове *учитель* нет приставки, слово *стол* состоит только из корня и нулевого окончания и не содержит ни суффиксов, ни приставок.

Наряду с выделенными типами морфем в сложных словах встречаются особые элементы, служащие для соединения частей этих слов (например, соединительные гласные: *о* в слове *людоед*, *у* в слове *двуглавый*). Такие элементы иногда называют **интерфиксами** (от латинских *inter* 'между' и *fixus* 'прикрепленный'). Являются ли интерфиксы морфемами? Это зависит от принятой точки зрения. Если считать выражаемую ими идею «соединения» полноценным означаемым, то интерфиксы по определению должны считаться морфемами — элементарными языковыми знаками. В противном случае интерфиксы следует признать незначащими «прокладками», появляющимися по чисто фонетическим причинам (в частности, для удобства произнесения). Тогда они не являются языковыми знаками, а следовательно, не могут быть отнесены к морфемам.



31. С морфемы какого вида может начинаться слово?

32. Морфемой какого вида может оканчиваться слово (не следует смешивать понятия «оканчиваться» и «быть окончанием»)?

33. Какое наибольшее число морфем каждого вида может быть в слове?

34. Из скольких звуков может состоять корень слова?

35. Какое максимальное число звуков в корне?

## 5. ВНУТРЕННЕЕ СТРОЕНИЕ И РАБОТА АЛГОРИТМА МОРФЕМНОГО АНАЛИЗА СЛОВА

Подобно тому как математическая теория помогает решению практических задач, формальная система морфемики может послужить на пользу практике, а именно — сделать процедуру разбора по составу более строгой, если не полностью алгоритмичной.

Для простоты мы построим алгоритм разбора слова, исключив из рассмотрения сложные слова.

Как вам должно быть известно из курса информатики и электронно-вычислительной техники, описание всякого алгоритма можно представить как состоящее из четырех частей: 1) описание исходных данных вместе с возможными связями между ними и интерпретацией; 2) описание основных действий (т. е. команд, или процедур); 3) указание последовательности этих действий; 4) описание возможных аномалий, т. е. ситуаций, для которых алгоритмом предусмотрены специфические процедуры. Исходные данные и требования к ним мы будем описывать, как это принято в программировании, отдельно от остальных трех частей, которые описываются нерасчлененно.

### 5.А. Исходные данные алгоритма

Грамматический словарь (в дальнейшем — Словарь 1). В этом Словаре каждое слово имеет помету, изменяется оно или нет и если да, то как именно, т. е. как оно склоняется или спрягается. Пользуясь этим Словарем, можно выделять окончания в любой словоформе (т. е. в произвольной форме любого слова из этого Словаря).

Словарь основ (в дальнейшем — Словарь 2). Этот Словарь более информативен. Он содержит: 1) толкования основ; 2) списки корневых гнезд основ.

Толкование основы — это некоторый формализованный текст, описывающий ее значение. Толкование содержит информацию, образована ли данная основа от какой-либо другой и если да, то от какой именно. На тексты толкований накладывается следующее формальное требование: толкования основ, образованных одним и тем же способом, должны быть представлены по одной схеме. Поясним это требование на примере. Пусть основа *болтун* имеет толкование 'человек, которому приписано свойство «болтать»', тогда

основа *ворчун* должна получить толкование 'человек, которому приписано свойство «ворчать» '. В этих толкованиях основы *болта-* и *ворча-* — это как раз те основы, от которых образованы *болтун* и *ворчун* соответственно. Основы сами должны быть толкуемыми единицами Словаря основ. Это еще одно формальное требование к структуре Словаря 2.

Что касается списка *корневых гнезд основ*, составляющего важный фрагмент данного Словаря, то о нем достаточно сказать следующее. Каждое корневое гнездо состоит из основ, входящих в данный Словарь и имеющих один и тот же корень. Так, в этот список входят гнезда {*дуб-*; *дубов-*; *дубин-*; *дубовик-*; ...}, {*свисте-*; *свист-*; *свистун-*; *свиста-*; *пересвист-*; *посвистыва-*; ...}, {*вчера-*; *вчерашн-*; *позавчера-*; *позавчерашн-*}.

Очень важное требование, накладываемое на оба Словаря, состоит в их согласованности. Словари 1 и 2 предполагаются не изолированными друг от друга: для каждой основы Словаря 2 в Словаре 1 должно найтись слово с этой основой. Аналогично для каждого слова Словаря 1 в Словаре 2 должна содержаться его основа.

Поскольку мы исключили из рассмотрения сложные слова, будем считать, что в Словаре 1 их нет, а тогда, в силу согласованности Словарей, в Словаре 2 нет сложных основ (основ, содержащих более одного корня).

К исходным данным относятся также:

*Список чередований*, применимых к морфам русского языка. Список чередований образуют все пары, в каждой из которых элементы могут заменять друг друга в результате применения операции чередования. В него входят пары *г~ж* (ср. *нога~ножка*), *ск~щ* (ср. *искать~ищейка*), но не входят пары *а~б* или *р~п*.

*Литерная переменная W*. *W* принимает свое значение на множестве словоформ всех единиц грамматического словаря.

*Литерные переменные S и B*. Названные переменные принимают свое значение на множестве всех единиц Словаря основ. Их содержание раскрывается в описании работы алгоритма.

Подведем итог. К исходным данным алгоритма относятся два Словаря, список чередований и три литерные переменные. Их необходимость станет ясна при разборе действия алгоритма.

## 5.В. Описание алгоритмических процедур

Допустим, мы хотим разобрать по составу некоторую словоформу. Условимся считать, что она подается на вход в

обычной буквенной записи, но буквы *е, ё, и, ю, я* после гласных, а также после букв *ѣ* и *ь* заменены на *йэ, йо, йи, йу, йа* соответственно.

Такая замена нужна, поскольку буквы *е, ё, и, ю, я* в указанных позициях передают, как известно, по два звука и часто бывает, что первый из них относится к одному морфу, а второй — к другому.

В начале действия алгоритма значением литерной переменной *W* является поданная на вход словоформа.

Перейдем к описанию этапов работы алгоритма.

**Выделение основы (этап 1).** Есть две логические возможности.

1) Поданная на вход словоформа представляет собой неизменяемое слово. Этот факт устанавливается по Словарю 1.

В случае неизменяемых слов их основа совпадает с самим словом, т. е., говоря алгоритмическим языком, *S* принимает значение *W* (*S* := *W*). Мы переходим сразу к этапу 2, чтобы установить, из каких морфов состоит основа.

**П р и м е р.** Пусть на вход подана словоформа *метро*. *S* принимает значение: *метро*.

2) Поданная на вход словоформа не совпадает ни с одним из неизменяемых слов Словаря 1.

Тогда, пользуясь информацией, содержащейся в Словаре 1, а) определяем, формой какого слова она является; б) выделяем в ней окончание (и тем самым — основу). Переходя на алгоритмический язык, можно сказать, что значением *S* в этом случае является та часть словоформы *W*, которая получится, если отбросить окончание.

**П р и м е р.** Пусть на вход подана словоформа *учительницей*. *S* принимает значение *учительниц*- (морф *-ей* признается окончанием).

Может случиться, однако, что поданная на вход словоформа не соотносима ни с одним из слов Словаря 1, т. е. не совпадает ни с одной из форм слов этого Словаря. Например, это может быть, когда на вход подается неологизм типа *прозаседавшиеся*, диалектизм типа *кочет* или форма сложного слова. Эта ситуация по определению считается аномальной. Алгоритм не работает.

**Выделение приставок и суффиксов (этап 2).** 1) Если основа *S* не образована ни от какой другой основы (как, например, основы слов *одеть* или *дом*, т. е. *оде-* и *дом-*), что устанавливается по Словарю 2, то переходим к этапу 3. 2) Если же *S* образована от другой основы *B*, то ищем максимальную совпадающую часть (м.с.ч.) основ *S* и *B*. При этом важно отметить, что совпадение здесь понимается с точностью до

чередований, т.е. не обязательно как полное графическое совпадение, но также, возможно, и как неполное: одна часть получается из другой в результате применения ряда операций чередования из исходного списка чередований.

Приведем несколько примеров таких совпадений. У основ  $S = \text{учительниц-}$  и  $B = \text{учитель м.с.ч.} = \text{учитель}$ , у основ  $S = \text{иррациональн-}$  и  $B = \text{рациональн- м.с.ч.} = \text{рациональн-}$ , у основ  $S = \text{подберезовик-}$  и  $B = \text{берез- м.с.ч.} = \text{берез-}$ , у основ  $S = \text{подружек}$  и  $B = \text{подруг- м.с.ч.} = \text{подруг-/подруж-}$  (чередование  $г \sim ж$ ), у основ  $S = \text{подоконник}$  и  $B = \text{окн- м.с.ч.} = \text{окн/окон}$  (чередование  $\emptyset \sim о$ ).

Часть, получаемая из  $S$  «вычитанием» максимальной совпадающей части, представляет собой, согласно данным ранее формальным определениям, либо приставку, либо суффикс, либо сложное образование «приставка — суффикс». Это устанавливается по местоположению максимальной совпадающей части в основе  $S$ . Например, в основе  $\text{учительниц-}$  выделяется суффикс  $\text{-ниц-}$ , в основе  $\text{подружк-}$  — суффикс  $\text{-к-}$ , в основе  $\text{иррациональн-}$  — приставка  $\text{ир-}$ , в основе  $\text{подстаканник-}$  (так же, как и в основе  $\text{подоконник-}$ ) — приставка  $\text{под-}$  и суффикс  $\text{-ник}$ .

После всех этих процедур алгоритм присваивает переменной  $S$  новое значение, а именно —  $B (S := B)$ . И к этому новому значению опять применяются процедуры этапа 2.

Посмотрим, что на этапе 2 произойдет с основой  $\text{учительниц-}$ . Алгоритм в основе  $\text{учительниц-}$  выделит часть  $\text{учитель-}$  и суффикс  $\text{-ниц-}$ . После этого переменной  $S$  алгоритм присвоит значение 'учитель'. Пользуясь Словарем 2 и переменной  $B$ , выделяем в основе  $\text{учитель-}$  часть  $\text{учи-}$  и суффикс  $\text{-тель}$ . Теперь значение  $S$  становится равным  $\text{учи-}$ . Основа  $\text{учи-}$  не образована ни от какой другой основы (это утверждение делается на основании Словаря 2). Поэтому переходим к этапу 3.

Как и для многих лингвистических алгоритмов, из-за недостаточной строгости ряда формулировок мы можем столкнуться с ситуацией неединственности результата. Пусть на вход этапа 2 поступает основа  $S = \text{вождений}$ . По Словарю основ определяем, что  $S$  образована от основы  $\text{води-}$  (поэтому  $B := \text{води-}$ ). На этапе 2, как мы помним, происходит выделение максимальной совпадающей части основ  $S$  и  $B$  — точнее, совпадающей с точностью до чередований. Поскольку список чередований, входящих в исходные данные нашего алгоритма, содержит чередования  $д \sim ж$  и  $д \sim жд$ , в принципе возникает неединственность дальнейшего членения. Максимальная совпадающая часть может быть выделена двумя разными способами:  $\text{вож-}$  (отличается от  $\text{вод-}$  на чередование  $д \sim ж$ )

и *возжд-* (отличается на чередование *д-жд*). Соответственно мы должны были бы выделить в первом случае суффикс *-дений*, во втором — *-ений*. Можно ли считать, что у нас два равноправных ответа? Ведь суффикс *-ений* и так выделяется, причем в большом числе основ (ср. *торможений* от *тормози-*, *служений* от *служи-*, *определений* от *определи-* и т. д.), а основ, где, бесспорно, выделялся бы суффикс *-дений* и нельзя было бы выделить *-ений*, совсем нет. Таким образом, разумно считать, что «задача» имеет одно решение: м.с.ч. основ *возждений* и *води-* — *возжд-*, а суффикс первой основы — *-ений*. Таким образом, теоретически возможную неединственность можно свести к практической единственности ответа в случае, если будут уточнены описания процедур и формулировки определений. В данном примере мы устраняем неединственность посредством сопоставления разбираемой основы с другими, имеющими в Словаре основ аналогичное толкование. Это позволяет сделать процедуру выделения максимально совпадающей части алгоритмизуемой.

**Выделение корня (этап 3).** На вход этапа 3 поступает основа *S*, не образованная ни от какой другой основы. Можно ли считать ее корнем? Нет, ее следует сначала сравнить с родственными основами.

Алгоритм обращается к Словарю 2, а именно к тому его фрагменту, где содержатся все корневые гнезда. Ищем корневое гнездо, в которое входит *S*. Оно обязательно найдется, т.е. алгоритм не будет работать безрезультатно. Корнем объявляется максимальная совпадающая часть всех элементов этого гнезда. Если максимальная совпадающая часть не тождественна *S*, то мы, помимо корня, выделяем в *S* приставку или суффикс.

На этом алгоритм заканчивает работу.

## 5.С. Примеры

Пусть *S* = *дом*. Основа *дом-* на этапе 3 сопоставляется с корневым гнездом {*домик*; *бездомн-*; *домашн-*; ...}. М.с.ч. = *дом*; следовательно, *дом* — корень.

Теперь пусть *S* = *учи-*. Основа *учи-* сопоставляется с гнездом {*учитель*; *ученик*; *выучк-*; *учебн-*; ...}. М.с.ч. = *уч-*; следовательно, в основе *учи-* часть *уч-* — корень, а *-и-* — суффикс.

В заключение приведем пример полного разбора словоформы *W* = *коснулся* шаг за шагом вслед за описанным алгоритмом.

На этапе 1 по грамматическому словарю узнаем, что *W* — это глагол в форме прошедшего времени единственного числа мужского рода словарной единицы *коснуться*. По тому

же словарю выделяем нулевое окончание со значением мужского рода единственного числа (ср. *коснулась, коснулось, коснулись*) и прерывистую основу *коснул..ся* ( $S = \text{коснул..ся}$ ).

На этапе 2, сравнивая  $S = \text{коснул..ся}$  и  $B = \text{косну..ся}$ , выделяем основу *косну..ся* и суффикс *-л-*.

На этапе 3, рассматривая корневое гнездо {*косну..ся; каса..ся; прикосновений;...*}, выделяем в  $S = \text{косну..ся}$  корень *кос-* и суффиксы *-ну-* и *-ся*. Алгоритм, следовательно, дает такое членение словоформы *коснулся*: *кос-* + *-ну-* + *-л-* +  $\emptyset$  + *-ся*.



36. Какое членение даст алгоритм для слов *переподготовка, впустую, признание, признательного, читательский, продумавший*?

## II. ПОНЯТИЯ И МОДЕЛИ СЛОВООБРАЗОВАНИЯ

### 1. ОТНОШЕНИЕ ПРОИЗВОДНОСТИ. БАЗОВОЕ И ПРОИЗВОДНОЕ СЛОВА

Часто при формальном описании слова удобно говорить, что одно слово образовано от другого слова или нескольких слов. Например, слово *выключатель* образовано от *выключать* (действительно, *выключатель* — это приспособление для включения и выключения механизмов и бытовых приборов); слово *домик* — от *дом* (*домик* — дом, который говорящий считает маленьким); *мыться* — от *мыть*; *лесостепь* — от слов *лес* и *степь* и т.д. В этих случаях между соответствующими словами устанавливают отношение производности: *выключатель* считается производным от *выключать*, *домик* от *дом*, *лесостепь* от пары <*лес; степь*> и т.д. Слова *выключать*, *дом* и пара <*лес; степь*> называются **б а з а м и** для своих производных. Все слова, входящие в базу, называются **б а з о в ы м и с л о в а м и** для своего производного.

Перечислим важнейшие свойства этого отношения.

1. Базовое и производное слова принадлежат одному корневому гнезду.

2. Основа производного слова, как правило, по составу морфем устроена сложнее, чем основа базового слова. Так, основа слова *выключатель* содержит те же морфемы, что и основа слова *выключать* (т.е. *выключа-*), плюс еще суффикс *-тель*.

3. В толкование производного слова принято включать базовое: *читальня* = 'место, отведенное для того, чтобы читать' (*читать* — базовое слово).

4. Из сказанного выше следует, что отношение производности не является симметричным: более того, если некоторое слово  $\beta$  признается производным от базового слова  $\alpha$ , то  $\alpha$



не может быть производным от  $\beta$  — ведь оно в паре  $\langle \alpha ; \beta \rangle$  является базовым.

5. Очевидно, что отношение производности не является рефлексивным.

6. Это отношение сознательно делают также нетранзитивным, то есть устраняют транзитивность: для каждого производного принято выделять ровно одну базу. Разумеется, это не препятствует тому, чтобы одно и то же слово было базовым для ряда производных: ср. пары *дом* — *домик*, *дом* — *домище*, *дом* — *домовой*, *дом* — *бездомный*.

Слово, которое не является производным ни от какой базы, называется *н е п р о и з в о д н ы м*. Непроизводными являются слова *дом*, *петь*, *старый*, *столб*.

В языке производных слов гораздо больше, чем непроизводных. Благодаря этому мы, зная одно слово, легко понимаем ряд других, производных от него. Слово *виноградина* встречается достаточно редко. Но даже впервые столкнувшись с ним, мы поймем, что оно значит 'одна штука винограда'; это мы устанавливаем по аналогии со словами *картофелина*, *горошина*, *жемчужина*, *льдина*, *рыбина* и др. Тем самым отношение производности помогает овладению лексикой языка.

## 2. СЛОВООБРАЗОВАТЕЛЬНАЯ ЦЕПОЧКА

Не следует думать, что все базовые слова являются непроизводными.

Базовое слово по отношению к некоторому слову может вместе с тем быть и производным от некоторой другой базы. Например, слово *учитель* по отношению к глаголу *учить* является производным. Оно означает 'человек, который учит', и в нем выделяется суффикс *-тель* со значением 'производитель действия'. А по отношению к слову *учительница* слово *учитель* является базой. *Учительница* — это 'женщина-учитель'.

Таким образом, мы получаем последовательность однокоренных слов, такую, что каждые два соседних слова находятся в отношении производности, причем первое из них является базовым, а второе — производным. Такая последовательность называется *с л о в о о б р а з о в а т е л ь н о й ц е п о ч к о й*.

Полная словообразовательная цепочка всегда начинается с непроизводного слова. Слово, производное от него, называется производным I степени, его производное — производным II степени и т.д. Таким образом, словообразовательная цепочка позволяет установить словообразовательную связь между словами, находящимися даже на противоположных, далеких друг от друга концах.

Примеры словообразовательных цепочек: *смелый — смельчак, дарить — дар, готовый — готовить — сготовить*.



37. Привести пример словообразовательной цепочки, в которую входит производное III степени.

38. Производным какой степени является существительное *устарелость*?

39. Постройте словообразовательные цепочки для слов *фронтоник, подготовительный*.

### 3. СЛОВООБРАЗОВАТЕЛЬНОЕ ГНЕЗДО

Существует еще более сложный механизм анализа словообразовательных связей. Речь идет о словообразовательном гнезде. Важность понятия словообразовательного гнезда определяется тем, что словообразовательное гнездо используется в качестве единиц в словообразовательных словарях, т. е. словарях, назначение которых — указать словообразовательные связи лексических единиц и их формальное строение.

С л о в о о б р а з о в а т е л ь н о е   г н е з д о — это множество словообразовательных цепочек, начинающихся с одного и того же слова. Словообразовательные гнезда изображаются в виде «деревьев».

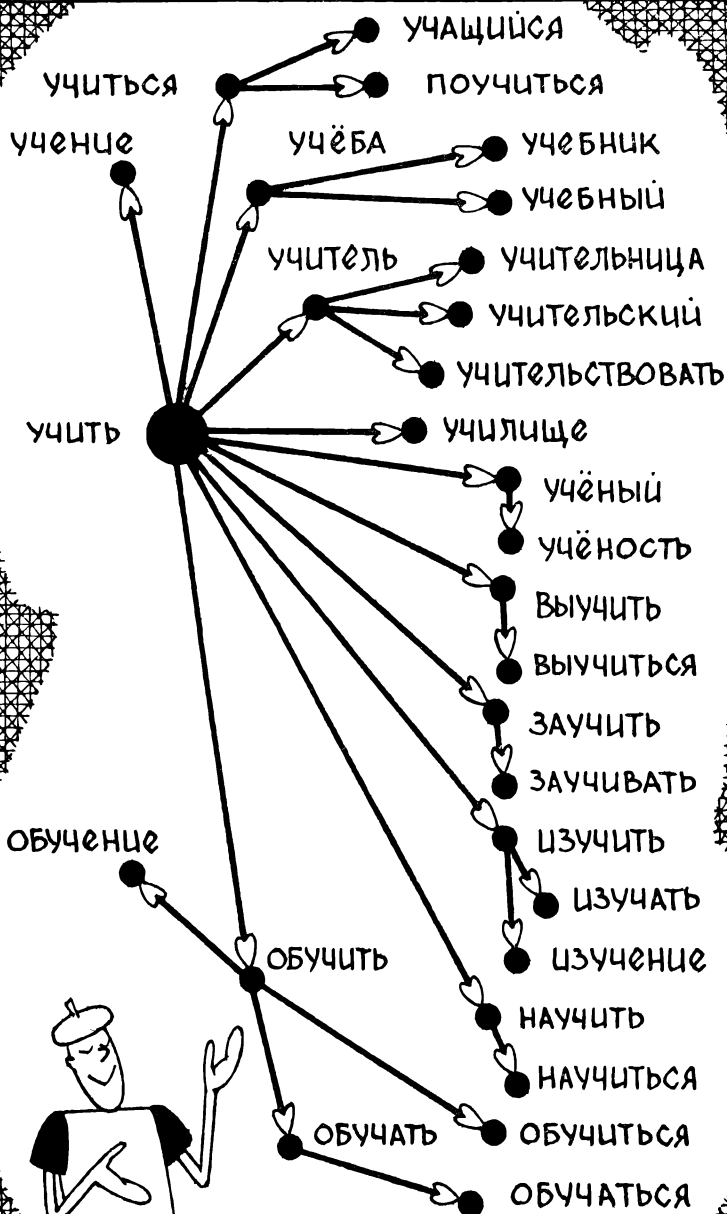
На рисунке на с.88 показан лишь один фрагмент словообразовательного гнезда, начинающегося словом *учить*.

С понятием словообразовательного гнезда связывают формальные признаки, передающие информацию о характере словообразовательных связей внутри гнезда: число «ветвлений» в словообразовательном гнезде — это число словообразовательных цепочек в составе гнезда «ветвления» (максимальная глубина «ветвления» — это длина (т. е. количество слов) максимальной цепочки) и др. Названные формальные признаки дают возможность производить количественную оценку и сравнение словообразовательных гнезд как внутри языка, так и в разных языках.

### 4. СЛОВООБРАЗОВАТЕЛЬНАЯ КАТЕГОРИЯ

#### 4.А. Словообразовательное значение

Как говорилось выше, производные слова толкуются при помощи базовых. Часть толкования, остающаяся после отбрасывания базового слова, передает с л о в о о б р а з о в а т е л ь н о е   з н а ч е н и е данного слова. Например, *читальня* имеет словообразовательное значение 'место, предназначенное для того, чтобы...'; такое же словообразовательное значение имеют слова *спальня, раздевалка, свалка*. Слово *стульчик*, как и слова *домик, столик, кроватка*, имеет словообразовательное значение 'маленький...'.



#### 4.В. Понятие словообразовательной категории

Словообразовательная категория — это множество слов, имеющих одно и то же словообразовательное значение. Например, слова *столлик, стульчик, кроватка, домик, человечек* и др. объединены общим словообразовательным значением 'маленький...' и тем самым принадлежат одной словообразовательной категории.



40. Даны глаголы, образованные от существительных или прилагательных: *озорничать, gloхнуть, голодать, лодырничать, пилить, звереть, свирепствовать, нуждаться, хамить, горевать, каменеть, рыбачить, блаженствовать, бороновать, толстеть, бинтовать, столбенеть, сверлить*.

Разбить эти глаголы на словообразовательные категории. Для каждой категории указать словообразовательное значение входящих в нее слов.

(Эта задача в несколько ином варианте предлагалась на III Олимпиаде по языковедению и математике, проводимой Московским государственным университетом для школьников в 1966/67 гг.)

Изучение словообразовательной категории позволяет выявить тонкие смысловые различия входящих в нее слов.

Рассмотрим словообразовательную категорию, объединенную значением 'тот, кто...', т. е. категорию «имен деятеля». Сюда входят слова *водитель, учитель, болтун, сварщик, победитель, советчик, командир* и многие другие. Между тем более внимательный анализ позволяет обнаружить четыре варианта данного значения. Эти варианты соотносятся с четырьмя логически возможными способами описания человека по производимому действию. Человек может характеризоваться по действию, которое он производит сейчас, в данный момент (*посетитель*), по действиям, которые ему свойственно совершать, хотя, быть может, сейчас он их не совершает (*болтун*), по действиям, которые ему положено совершать, так сказать, по статусу (*спасатель* на лодочной станции), по действиям, которые он хотя бы раз совершил в прошлом (*спаситель*). Интересно, что обычно одно и то же слово может выступать в нескольких вариантах данного значения: например, слово *водитель* может означать человека, который сейчас ведет транспортное средство (*Не отвлекайте водителя!*), а также человека, которому положено водить в силу его профессиональных обязанностей (*Все водители вышли на линию*). Именно большое количество подобных случаев неединственности допустимого осмысления обеспечивает цельность словообразовательной категории «имен деятеля».

#### 4.С. Словообразовательный тип

Внутри одной словообразовательной категории выделяют подмножества — словообразовательные типы. Словообразовательный тип — это множество слов, входящих в одну словообразовательную категорию (т.е. имеющих одно словообразовательное значение) и образованных при помощи одного и того же формального средства. В частности, внутри словообразовательной категории «уменьшительных существительных» выделяются типы «слова с суффиксом *-чик*» (*стульчик; вагончик; ...*), «слова с суффиксом *-ик*» (*столик; прыщик; домик; ...*), «слова с суффиксом *-ок*» (*лесок; ветерок; ...*) и т.д.

Если вам удалось решить предыдущую задачу, то вы легко справитесь со следующей.



41 (продолжение задачи 40). Каждая из выделенных вами словообразовательных категорий состоит из словообразовательных типов. Указать все эти типы и для каждого из них назвать соответствующее словообразовательное средство.

#### 4.Д. Пример формального описания словообразовательной категории

Почему жителей Волгограда зовут *волгоградцы*, а жителей Киева — *киевляне* (а не «*киевцы*»)?

Образование названия жителей от названия населенного пункта представляет собой сложную и очень важную лингвистическую проблему. Ведь одно и то же содержание здесь выражается разными способами. Для образования названий жителей используются такие суффиксы, как *-ц-* (*волгоградцы*), *-ан-* (*калужане*), *-чан-* (*ростовчане*), *-ич-* (*москвичи*) и др. Какие же обстоятельства влияют на столь сложный выбор правильной формы?

Во-первых, это принадлежность названия жителей к классу разговорных или официальных названий. Например, жители Пошехонья официально называются *пошехонцы*, а разговорное их название — *пошехоны*. Очевидно, что разговорные названия жителей гораздо более разнообразны, чем официальные, и строятся по своим, весьма прихотливым законам.

Поэтому формальная модель строится лишь для официальных названий.

Между тем даже не все официальные названия описываются формальной моделью. Есть населенные пункты, жители которых называются вопреки формальным правилам. Это в первую очередь связано с тем, что данное название закреплено в истории города или в национальной традиции. Например, жители Смоленска называются *смоляне*, в Тверской области есть город Осташков; его жителей издавна зовут

*осташи*. В Туле живут *туляки*, в Курске — *куряне*, в Одессе — *одесситы*, а в Ржеве — *ржевитяне*. А как бы вы думали, называют жителей Архангельска? Если человек не знает, то он ни за что не догадается. Ведь это название образуется не по общим правилам и звучит так: *архангелогородцы*! А жителей Торжка — *новоторы* или *новоторжцы*. Все названия жителей, образованные при помощи суффикса *-ич-*: *москвичи*, *псковичи*, *костромичи*, *томичи* (жители Томска), *омичи* (жители Омска), *серпуховичи* — также не порождаются правилами. Все эти наименования, относящиеся к исключениям, как и другие исключения, приходится заучивать наизусть.

Интересно, что есть города, жители которых могут называться не единственным способом. Например, жители Ставрополя могут называться *ставропольцы* и *ставропольчане*, а жители города Спасска — *спасцы*, *спасовцы* и *спасчане*. При этом одна из форм, возможно, употребляется чаще, чем другая, но это определяется статистическими подсчетами и непосредственно с правилами словообразования не связано.

Таким образом, словообразовательным значением 'люди, которые живут в...' обладают производные слова, образованные от базовых слов — названий географических пунктов. Если исключить из рассмотрения те нестандартные способы образования названий жителей, о которых мы только что говорили, то в данной словообразовательной категории можно выделить семь основных словообразовательных типов. (Все нижеследующие примеры см.: Словарь названий жителей РСФСР/ Под ред. Н.М.Бабкина. — М., 1964.) Каждый тип характеризуется своим суффиксом, присоединяемым к основе базы, а также теми изменениями, которыми такое присоединение может сопровождаться.

При построении производных слов говорящие подсознательно стремятся сохранить по возможности больше сходств между базовым и производным словами. Понятно, что в паре *Таруса* — *тарусцы* основы различаются меньше, чем в паре *Таруса* — *тарусяне* или в паре *Таруса* — *тарушане*.

Идея минимизации числа несовпадающих признаков «парных» слов и лежит в основе лингвистического алгоритма построения названия жителей по заданному названию места (сам алгоритм мы здесь не приводим). Такой алгоритм породит форму *тарусцы* и забракует формы *тарусяне* или *тарушане*.

Чтобы выделить те признаки основ, которые надо учитывать при построении алгоритма, лингвист сначала подробно описывает каждый из семи основных способов образования названий жителей. Он указывает суффиксы, задающие эти способы: *-ц-*, *-чан-*, *-ан-* (*-ян-*), *-овц-* (*-евц-*), *-инц-*, *-анц-* (*-янц-*),

-енц-, и определяет то, как они воздействуют на основу. Например, суффикс *-чан-* всегда «перетягивает» на себя ударение (*Ростóв — ростовча́не*). Суффикс *-ц-* «не любит» согласных *г, к, х* в конце основы базового слова — названия города: поэтому его присоединение к основе, оканчивающейся на данные согласные, нередко приводит к чередованию (например, *Выбо́рг — выбо́ржцы*).

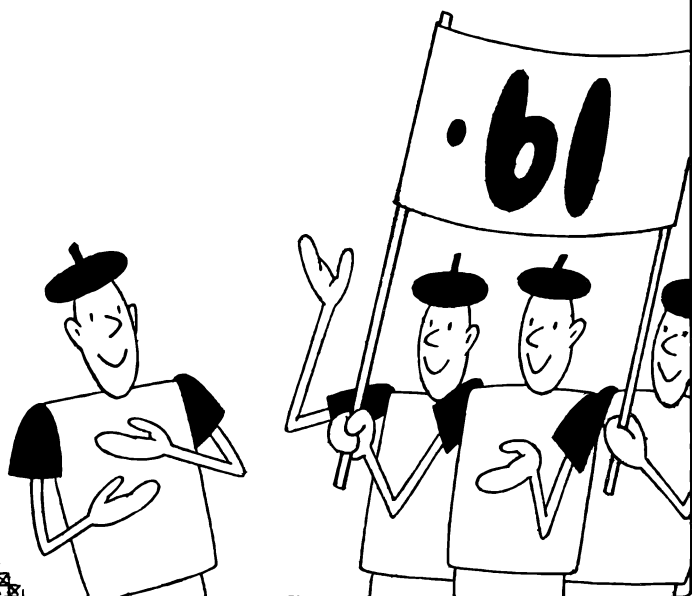
Именно исчисление всех возможностей подобного рода и позволяет выделить те признаки исходного слова, на базе которых строится алгоритм. К ним относятся, в частности: позиция ударения (принимающий значения «на основу» — «не на основу»), признаки согласного на конце основы («шипящий», «мягкий» и др.), количество согласных на конце основы («одиночный» — «стечение двух и более согласных»).

Отметим, что при минимизации различий алгоритм учитывает не только сами признаки, но и их «относительный вес», т.е. большую или меньшую значимость признака по сравнению с другими. Тем самым в алгоритм привносится количественная оценка изменений, к которым приводит присоединение данного суффикса к данной основе: чем число изменений меньше, тем более вероятным будет образование названия жителей от данного названия места именно при помощи данного суффикса.

Свойство алгоритма опираться на преобладание тенденции (выраженное количественно), а не на жесткие правила, учет не только качественных, но и количественных соотношений — все это, естественно, ведет к большой сложности и большому объему алгоритма. Поэтому мы ограничились здесь описанием одной только идеи его построения. Попытка подробного анализа структуры такого алгоритма (впрочем, также без самого алгоритма) дана в работе: А р а п о в М. В. Соотношение существительных со значениями «название местности» и «название жителей данной местности» в современном русском языке // Проблемы структурной лингвистики. 1971. — М., 1972. Хотим предупредить читателя, мало знакомого с применением вероятностных моделей в формальной лингвистике, что эта работа, достаточно сложная даже для специалистов, будет ему, скорее всего, непонятна.

Глава пятая

**Грамматические  
категории и их  
моделирование**





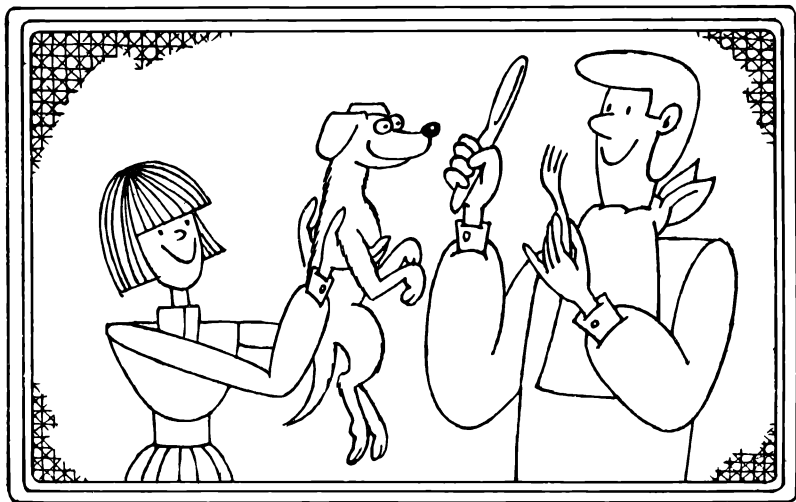
# 1. ГРАММАТИЧЕСКАЯ ИНФОРМАЦИЯ О СЛОВЕ

## 1. ПРАВИЛА ПОВЕДЕНИЯ СЛОВ

Слова языка подобны живым организмам. Они имеют внешний облик, внутреннюю ценность, или значимость, в системе и обладают особыми характеристиками поведения. Для того чтобы правильно пользоваться словами, надо знать не только, как они произносятся и что они значат, но и как им положено вести себя в речи. Информация о правилах поведения слов называется грамматической информацией.

Возьмем, например, слово *накормить*. Мы знаем, что правильно сказать *Я накормила собаку*, но не *Я накормила собаке*. Между тем говорят *Я дала поесть собаке*, но нельзя сказать, по крайней мере в естественной ситуации: *Я дала поесть собаку*, хотя очевидно, что *накормить* означает почти то же, что и *дать поесть*.

Это свойство глагола *накормить* можно описать так: глагол *накормить* требует винительного, а не дательного падежа дополнения. Для выражения будущего времени этого глагола мы используем простую, а не сложную форму: *Я накормлю*, а не *буду накормить*. Не располагая подобной информацией, мы не построим правильных русских предложений со словом *накормить*.



## 2. ПОНЯТИЕ ГРАММАТИЧЕСКИХ КАТЕГОРИЙ

В лингвистике информацию о поведении слова принято описывать с помощью понятия 'грамматическая категория'. В школе вы уже сталкивались с такими грамматическими категориями, как род, число, падеж существительного, вид и время глагола. Г р а м м а т и ч е с к а я к а т е г о р и я — это частный случай общенаучного понятия признака, и поэтому говорят о значениях грамматической категории как о значениях признака. Например, грамматическая категория «род существительного» имеет значения «мужской род», «женский род», «средний род».

### 2.А. Типы грамматических категорий

Грамматические категории бывают двух типов: классифицирующие и словоизменятельные.

К л а с с и ф и ц и р у ю щ а я к а т е г о р и я — это признак, принимающий одно и то же значение у всех форм, относящихся к одному слову. Например, род существительного — классифицирующая категория. Действительно, если существительное *стол* в каком-то падеже относится к мужскому роду, то и во всех других падежах это слово мужского рода.

В силу такого понимания классифицирующей категории можно говорить, что ее значение характеризует слово как единицу словаря.

С л о в о и з м е н и т е л ь н а я к а т е г о р и я — это признак, принимающий разные значения хотя бы у двух форм одного слова. Например, число существительного — словоизменятельная категория: форма *стол* — единственного числа, а форма *столы* имеет значение множественного числа. Заметим, что род прилагательного, в отличие от рода существительного, — словоизменятельная категория: *красный* — форма мужского рода, а *красная* — женского.

Каждое слово в словаре характеризуется набором значений классифицирующих категорий, но ему нельзя приписать набор значений словоизменятельных категорий, поскольку эти значения, вообще говоря, меняются от формы к форме. Однако важной грамматической характеристикой слова в словаре является множество самих словоизменятельных категорий, по которым оно изменяется. Говоря формально, грамматическое поведение каждого слова моделируется с помощью кортежа, состоящего из разнородных элементов: значений классифицирующих категорий и названий категорий словоизменятельных.

## 2.В. Понятие грамматического разряда слова

Множество слов, характеризующихся одним и тем же набором словоизменительных категорий, называется **г р а м м а т и ч е с к и м   р а з р я д о м**. Например, все существительные, изменяющиеся по числам и падежам, образуют один грамматический разряд слов. Слова *один, два* и *три* относятся к разным грамматическим разрядам. Действительно, *один* изменяется по числам, родам и падежам, *два* — по родам и падежам, а *три* — только по падежам.



42. Разбить следующие слова на грамматические разряды: *полтора, сотня, сто, два, сотый, кто, какой, несколько, оба, двадцать первый*.

43. Для каждого грамматического разряда, выделенного в предыдущей задаче, указать еще одно-два слова, входящие в этот разряд.

Ниже мы покажем, как можно моделировать грамматическое поведение слова. Грамматическое поведение слова имеет две стороны. В одном случае оно, так сказать, пассивное: мы должны знать, какую форму принимает данное слово под воздействием окружения, или контекста. Здесь контекст является активным началом, а поведение слова подчинено ему. В другом случае активно уже само поведение слова. Здесь слово диктует свою волю соседним словам. В соответствии с этими особенностями моделирование приобретает различный характер. Сначала мы рассмотрим, как строится логическая конструкция, назначение которой — формально представить черты поведения первого типа. Далее продемонстрируем, как другая логическая конструкция служит моделью поведения второго типа.

## II. ВНУТРЕННЯЯ СТРУКТУРА ГРАММАТИЧЕСКОЙ КАТЕГОРИИ: ЛОГИЧЕСКАЯ КОНСТРУКЦИЯ ПАДЕЖА

### 1. МОДЕЛЬ ПАДЕЖА СУЩЕСТВИТЕЛЬНЫХ В РУССКОМ ЯЗЫКЕ

Почему в русском языке у существительных обычно выделяют шесть падежей, а не, скажем, четыре или десять? Почему словоформы *мелом* и *лопатой* существительных *мел* и *лопата* считаются формами одного (творительного) падежа, хотя окончания в них разные, в то время как формы *сад<sub>1</sub>* в сочетании *стоит сад* и *сад<sub>2</sub>* в сочетании *вижу сад* признаются омонимичными формами разных падежей? Каковы вообще логические основания для выделения падежей и критерии их различения? На эти вопросы поможет ответить рассматриваемая ниже модель падежа существительного, представляющая

собой упрощенный вариант логической конструкции, предложенной известными учеными — математиком В.А.Успенским и лингвистом А.А.Зализняком.

## 1. А. Ограничения на модель

Не желая утомлять читателя слишком формальными построениями, введем некоторые ограничения на модель.

1. Из разнообразных аспектов пассивного поведения слова мы остановимся только на трех, которые имеют отношение к модели изменения существительных.

2. Модель будет построена для «предметных» имен существительных — существительных, обозначающих живые существа или конкретные предметы, т.е. таких слов, как *палка, друг, пальто, собака, ...*

3. Как известно, одни формы предметных существительных относятся к единственному числу, а другие — к множественному. В качестве основной единицы моделирования выбирается существительное с фиксированным значением числа, т.е. единица, рассматриваемая как множество форм, имеющих одно и то же числовое значение. Так, {*палке; палку; палкой;...*} считается одной единицей, обозначающей единичный объект «палка», а {*палкам; палками;...*} — другой единицей, обозначающей множественный объект «палки» (т.е. число у нас будет считаться как бы классифицирующей категорией — разные числа характеризуют разные объекты).

## 1. В. «Фраза с многоточием»

Теперь, после указанных ограничений на модель, введем понятие «фраза с многоточием». «Фраза с многоточием» — это последовательность, состоящая из слов и одного многоточия. Будем говорить, что данная фраза с многоточием допускает подстановку формы слова, или словоформы *L*, если результатом подстановки *L* на место многоточия является правильная фраза. С помощью операции подстановки словоформы во фразу с многоточием и определяется формально категория падежа.

Приводимое ниже описание иллюстрирует один из возможных путей такого определения.

### 1.С. Падежная таблица и операции над ней

Предположим, что у нас есть бесконечная таблица следующего вида. Сверху по горизонтали выписан полный список основных единиц моделирования. Слева по вертикали дан полный список фраз с многоточиями (именно последнее предположение делает таблицу бесконечной). Внутренняя часть таблицы заполняется таким образом: в клетке на

пересечении строки и столбца выписаны словоформы, принадлежащие данной единице и допустимые при ее подстановке в данную фразу с многоточием. В том случае, если таких словоформ нет, в соответствующей клетке ставится прочерк.

Таблица имеет следующий вид (разумеется, мы можем привести здесь лишь ее фрагмент):

	«палка»	«палки»	«друг»	«друзья»	«пальто» (одно)	«пальто» (много)
1. Я вижу...	палку	палки	друга	друзей	пальто	пальто
2. Я вижу свою...	палку	—	—	—	—	—
3. Я вижу свои...	—	палки	—	—	—	пальто
4. Я вижу своего...	—	—	друга	—	—	—
5. У меня нет...	палки	палок	друга	друзей	пальто	пальто
6. Я не нашел...	палки	палок				
7. Я дорожу...	палку	палки	друга	друзей	пальто	пальто
8. Мне нравится...	палкой	палками	другом	друзьями	пальто	пальто
9. Я люблю...	палка	—	друг	—	пальто	—
	палку	палки	друга	друзей	пальто	пальто

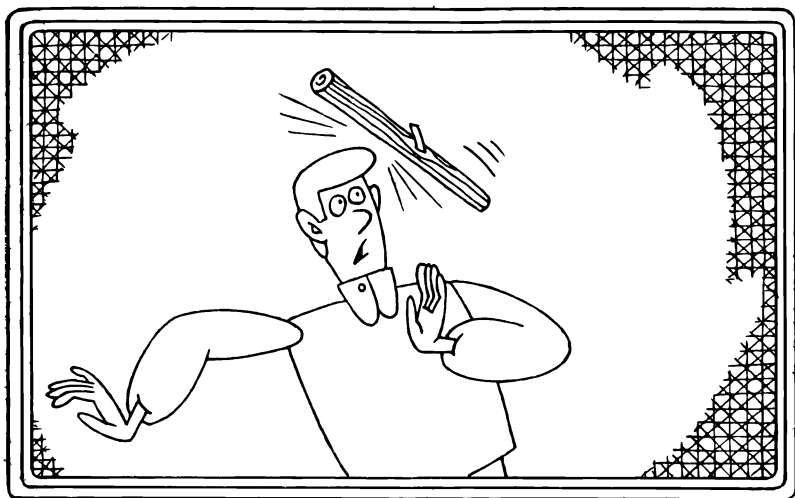
Пояснение: слово *L* в кавычках обозначает основную единицу моделирования, имеющую значение «*L*», т.е. «палка» = {палке; палку; палкой;...}, а «друзья» = {друзья; друзей; друзьям;...}.

Для формального определения падежа над данной таблицей производятся следующие операции в том порядке, в каком они ниже описываются.

1. Из всех строк, совпадающих внутри таблицы, оставляют ровно одну, а все остальные вычеркиваются. Так, в той части бесконечной таблицы, которую мы привели, совпадают первая и девятая строки, и одна из них (безразлично какая) вычеркивается.

2. Далее вычеркивается любая строка, содержащая прочерк, в том случае, если существует другая строка, совпадающая с ней в непустой части (т.е. в части, не содержащей прочерков). Например, в нашем примере таблицы придется, как нетрудно видеть, вычеркнуть 2, 3, 4-ю строки.

Оказывается, что в результате этой операции на практике вычеркиваются все строки, содержащие прочерк хотя бы в одной клетке.



3. Пусть проведены первые две операции. Чтобы определить третью операцию, нам понадобятся понятия «объединение клеток» и «объединение строк». Клетка  $a$  называется объединением клеток  $a_1, a_2, \dots, a_n$  той же таблицы, если содержимое  $a$  есть объединение содержимого клеток  $a_1, a_2, \dots, a_n$ . Строка  $A$  называется объединением строк  $A_1, A_2, \dots, A_n$ , если каждая клетка из строк  $A$  есть объединение клеток из строк  $A_1, A_2, \dots, A_n$ , принадлежащих тому же столбцу, что и эта клетка.

Третья операция над таблицей состоит в вычеркивании из нее всех строк-объединений. Например, будет вычеркнутой 6-я строка, являющаяся объединением 1-й и 5-й. В то же время 1-ю строку нельзя считать объединением 2, 3 и 9-й, поскольку к началу третьей операции 2-я и 3-я строки исчезнут.

Проделав эти три операции над исходной таблицей, мы получим новую таблицу с тем же числом столбцов, но с меньшим числом строк. Каждая строка такой таблицы называется однопадежным рядом. Всем словоформам в одном однопадежном ряду приписывается одно и то же значение категории падежа. В этом случае удобно говорить, что все эти словоформы стоят в одном и том же падеже или относятся к одному и тому же падежу. Разным однопадежным рядам соответствуют разные падежи.

## 2. ПРОБЛЕМА КОЛИЧЕСТВА ПАДЕЖЕЙ У СУЩЕСТВИТЕЛЬНОГО В РУССКОМ ЯЗЫКЕ

Сколько же падежей у существительного в русском языке? Теперь мы можем не только уточнить этот не вполне ясный и нестрогий поставленный вопрос, но и при соответствующих уточнениях и ограничениях дать на него простой и вместе с тем формальный ответ, а именно: падежей столько, сколько однопадежных рядов окажется в результирующей таблице.

В итоге в русском языке выделяются шесть общепризнанных падежей, а также по крайней мере еще один падеж. Дело в том, что среди однопадежных рядов обнаруживается строка типа: *Что висит на... (палке, палках, друге, друзьях, пальто, шкафу, печи, ведре...)?*

В самом деле, словоформы, произносимые как *шкафу́*, *печи́*, ни в каком однопадежном ряду больше не встречаются. Специально подчеркнем отличие этого ряда от ряда, соответствующего предложному падежу. В последнем, задаваемом, например, фразой *Я говорю о...*, в столбце «шкаф» будет форма *шка́фе*, а не *шкафу́*; *печи́*, а не *печи́*. Это обстоятельство заставляет нас выделить особый м е с т н ы й падеж, отличный от предложного.

Читатель, по-видимому, уже понял, что ответ на вопрос о количестве падежей в русском языке зависит от той формальной конструкции, которая положена в основу модели падежа. В другой модели число падежей могло бы оказаться иным.

Как бы то ни было, при выборе модели мы должны обеспечить выполнение следующего условия. Всякая модель падежа должна обладать двумя свойствами:

1. Любые два падежа, различаемые этой моделью, должны быть представлены внешне различными словоформами хотя бы у одной единицы: ведь иначе не было бы смысла противопоставлять один из этих падежей другому.

2. Эта модель должна формулировать правила, которые по двум объектам — единице моделирования и падежу — строят третий объект, а именно конкретную словоформу данной единицы, стоящую в данном падеже (возможно, в нескольких, различающихся по стилю вариантах, как в творительном падеже — *палкой* и *палкою*).

В силу последнего требования число падежей не может быть меньше пяти. В противном случае мы бы столкнулись с такой ситуацией: на входе правил мы имели бы единицу «палка», падежей у нас меньше пяти — и, следовательно, по крайней мере одной из единиц *палка*, *палки*, *палке*, *палку*, *палкой* мы не получим на выходе.

Построенная логическая конструкция падежа — это

всего лишь один пример моделирования грамматических категорий. Подобные конструкции формальная лингвистика умеет строить, во-первых, для других категорий, а во-вторых, для других частей речи. Целью грамматического моделирования является уточнение интуитивных понятий, стоящих за содержательными грамматическими категориями. Тем самым, все хорошие модели, которые описывают примерно один и тот же круг грамматических явлений, отличаются лишь частностями, больше говорящими о свойствах самой конструкции, чем о содержании понятия. В этом смысле можно считать, что хорошая модель грамматической категории всегда дает интуитивно приемлемый результат.

### III. ГРАММАТИЧЕСКИЕ КАТЕГОРИИ И СОГЛАСОВАНИЕ

#### 1. ВЛИЯНИЕ КАТЕГОРИИ РОДА НА ФОРМЫ ОКРУЖАЮЩИХ СЛОВ

Среди грамматических категорий имени существительного можно выделить по крайней мере одну, которая определяет прежде всего активные свойства данного имени, т.е. его влияние на формы окружающих слов. Речь идет о категории рода. В русском языке этот признак имеет три значения: мужской, женский и средний. Что же означает утверждение: «Слово *зима* женского рода, а *лето* — среднего»? Ответ на этот вопрос извлекается из поведения слов *зима* и *лето* в тексте. Слова, согласующиеся со словом *зима*, из множества возможных форм выбирают формы одного типа (*теплая зима, теплую зиму, теплой зимой*), а слова, согласующиеся со словом *лето*, — формы другого типа (*теплое лето, теплым летом*).

Род действует не только на непосредственно окружающие слова, но и на слова не из ближайшего окружения: во фразе *Первым пришел мой лучший друг* мужской род существительного *друг* диктует выбор формы не только слова *лучший*, но и формы слов *мой, пришел* и даже *первым* (ср. *Первой пришла моя лучшая подруга*).

Казалось бы, род существительного проявляет себя лишь когда оно стоит в единственном числе (ср.: *Уедут мои друзья* и *Уедут мои подруги*). Однако рассмотрим словосочетания *один из моих друзей* и *одна из моих подруг*. Выбор форм *один* (мужской род) и *одна* (женский род) производится в зависимости от рода существительных *друг* и *подруга*, хотя эти слова здесь и стоят во множественном числе. Таким образом, утверждение, что роды существительных во множественном числе не различаются, неверно, если придать этому утверждению следующий смысл: в любой правильной русской фразе с



именем существительным во множественном числе можно заменить это имя на существительное другого рода, поставив последнее в том же числе и падеже и получив при этом правильную фразу. Приведенный выше пример показывает, что такая замена не всегда возможна.

## 2. СОГЛАСОВАТЕЛЬНЫЕ КЛАССЫ

Поскольку род существительного проявляется в формах согласующихся с ним слов, то при построении формальной модели рода последнего всегда ставится в соответствие множество грамматических форм согласующихся с данным существительным слов. Например, мужскому роду ставятся в соответствие формы *новый, этот, мой, ...* в именительном падеже единственного числа, *нового, этого, моего, ...* — в родительном и т. д., а женскому роду — *новая, эта, моя, ...* — в именительном падеже, *новой, этой, моей, ...* — в родительном падеже и т. д.

### 2. А. Строение и количество согласовательных классов

Предположим, что каждому существительному мы поставили в соответствие такой набор форм. Если два существительных имеют одинаковые наборы в каждом из падежей, то говорят, что они входят в один согласовательный класс. Очевидно, что отношение «входить в один согласовательный класс» есть отношение эквивалентности. В силу теоремы о разбиении (см. первую главу) все существительные делятся на непересекающиеся согласовательные классы.

Построенная логическая конструкция в предположении, что она моделирует род русского существительного и при этом является хорошей, должна была бы отвечать интуитивным представлениям о роде существительных. Иными словами, должны были бы выделиться три согласовательных класса в соответствии с тремя родами. А теперь посмотрим, дает ли наша конструкция на самом деле этот результат.

Возьмем шесть существительных (два мужского рода, два женского и два среднего): *дом, друг, стена, подруга, окно, животное*. Рассмотрим поведение какого-нибудь прилагательного, сочетающегося со всеми этими словами (например, *новый*) в именительном падеже единственного числа и в винительном множественного.

	дом	друг	стена	подруга	окно	животное
И.ед.	новый	новый	новая	новая	новое	новое
В.мн.	новые	новых	новые	новых	новые	новых

Трех классов, как видим, не получается. Формы винительного множественного у прилагательных, согласующихся с членами пар *дом — друг, стена — подруга, окно — животное*, различаются. «Хитрость» довольно простая. Вторые члены каждой из этих пар одушевленные, а первые — неодушевленные. Иными словами, при моделировании одной категории — категории рода существительных — приходится учитывать влияние другой — категории одушевленности. Нам остается одно из двух: сохранив модель, сказать, что в русском языке шесть родов (поскольку выделяются шесть согласовательных классов), либо изменить модель, сохранив интуитивное представление о трех родах. В видоизмененной модели каждому роду соответствовал бы не один согласовательный класс, а два: одушевленный и неодушевленный. Хотя в таблице указаны только две строчки, задающие падежно-числовые формы, учет всех остальных возможных строк не изменит ни количество, ни состав классов. Читатель может это сам проверить — ведь строк в таблице конечное число.

Ну что, задача решена? Модель построена? К сожалению, русский язык не так прост, как могло бы показаться. Что нам делать, например, со словом *староста*? Ведь обе следующие фразы грамматически правильны: *У нас будет новый староста — Петя* и *У нас будет новая староста — Катя*. Если считать, что *староста* — это одно слово, то получится, что у нас не шесть согласовательных классов, а больше — ведь слово *староста* объединяет в себе свойства двух согласовательных классов: второго и четвертого. Чтобы не вводить еще каких-то новых согласовательных классов, можно считать, что есть два слова *староста* — одно 2-го класса, а другое — 4-го. Так же можно поступить и со словами *калека, сирота, зануда* и т. д., т. е. со всеми словами так называемого «общего рода». Тем самым наша формальная модель избавляется от расплывчатого понятия «общий род», не соответствующего никакому согласовательному классу, и сохраняет прежнее разбиение на классы.

Впрочем, и на этом трудности, связанные с моделью, не заканчиваются. Раз множество согласовательных классов есть некоторое разбиение множества русских существительных, то каждое из этих существительных должно попасть в один и только один класс. Куда же тогда попадут слова *сани, ворота, чернила, часы, сутки* и др., т. е. те слова, которые не употребляются в единственном числе? Остается надеяться, что род этих слов удастся опознать, помещая их в контексты типа *один из этих...*; *одна из этих...*, о которых шла речь выше. Посмотрим, что получится. Возьмем, например, слово *сани*. Нельзя порусски сказать ни *один из этих саней*, ни *одна из этих саней*, ни *одно из этих саней* (можно лишь *одни из этих саней*). Иными

словами, наш диагностический контекст не различает рода слова *сани*: оно не относится ни к мужскому, ни к женскому, ни к среднему роду и, следовательно, ни к одному из шести классов. Здесь волей-неволей приходится выделять седьмой согласовательный класс. Что же касается отнесения таких слов, как *сани*, к какому-либо роду, то можно поступить одним из двух способов: 1) не приписывать им никакого рода, в результате чего существительные русского языка не будут «без остатка» разбиваться на роды; 2) ввести для таких слов особый род, т. е. иметь в русском языке более трех родов.

## 2. В. Возможности распределения существительных по родам и согласовательным классам

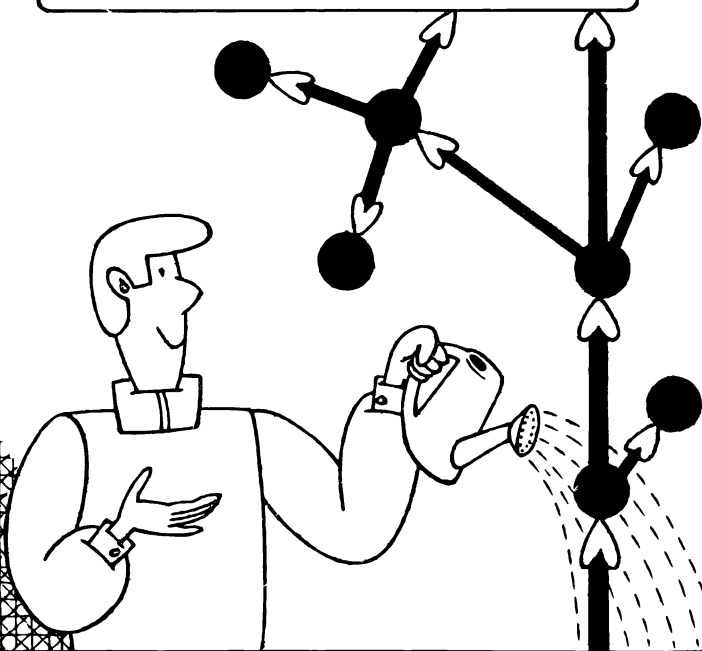
Итак, мы рассмотрели следующие логические возможности для имен существительных: одушевленные имена, имеющие обе числовые формы; неодушевленные имена, имеющие обе числовые формы; неодушевленные имена, имеющие только форму множественного числа. Остаются еще три логические возможности: одушевленные существительные, употребляющиеся только в единственном числе; неодушевленные существительные, употребляющиеся только в единственном числе; одушевленные, употребляющиеся только во множественном числе. Для первых трех случаев мы получили семь классов и три (или четыре?) рода. Одушевленных существительных, употребляющихся только в единственном числе, по данным грамматических словарей в русском языке нет (за исключением слова *Господь*). Род неодушевленных существительных, употребляющихся только в единственном числе, определяется так же, как и род многих неодушевленных существительных: *лай* — мужского, *медь* — женского, *золото* — среднего. Поскольку они не имеют множественного числа, их распределение по согласовательным классам условно: чтобы не увеличивать без особой на то необходимости число классов, их относят к 1, 3 и 5-му классу соответственно.

Остается последний случай: одушевленные существительные, употребляющиеся только во множественном числе. Это слова типа *ребята*, *девчата* и др. Снова применяя к ним контекст *один (одна) из этих...*, мы увидим, что каждое из них падает в один из уже выделенных нами классов.

Заключая эту часть главы, еще раз подчеркнем, что на синтаксическое поведение слов каждая категория оказывает влияние не сама по себе, но во взаимодействии с некоторыми другими категориями. В согласовании, например, принимают участие как классифицирующие категории существительных (род, одушевленность), так и словоизменительные (число, падеж).

Глава шестая

# Синтаксические законы



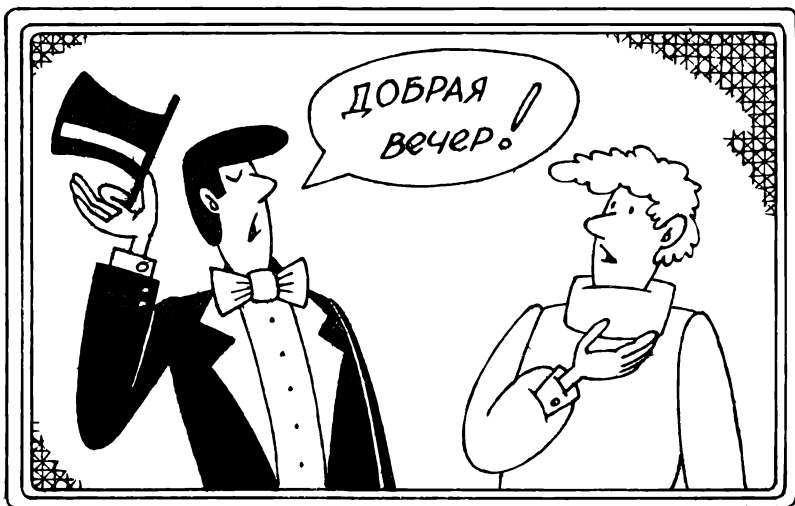
## 1. ФОРМАЛЬНЫЕ СИНТАКСИЧЕСКИЕ ПОНЯТИЯ

До сих пор мы рассматривали отдельные слова, но человек не говорит отдельными словами. Он связывает их в более крупные единицы: словосочетания и предложения. Даже самые простые мысли трудно передавать с помощью отдельных слов.

Между тем крупные единицы строятся из простых по своим законам. Свод таких законов носит название синтаксис.

### 1. ТИПЫ СИНТАКСИЧЕСКИХ ПРАВИЛ

Синтаксические правила бывают нескольких типов. Одни из них говорят о том, какой облик должны принимать слова, чтобы их можно было бы соединить друг с другом. Все мы знаем, что по-русски не говорят: *добрая вечер, шел леса, долг над родиной*, а надо сказать: *добрый вечер, шел лесом, долг перед родиной*. Такие правила называются правилами синтаксической связи. Именно эти правила нарушены в неправильных фразах *Она — большой умница* и *Его осудили к трем годам тюремного заключения*. Другие правила определяют правильную расстановку слов в словосочетаниях или предложениях. Если мы хотим выразить мысль, что Максим и Нина поступили в пищевой институт, мы можем это сделать, например, с помощью следующих предложений:



*Максим и Нина поступили в пищевой институт, Нина и Максим поступили в пищевой институт и даже В пищевой институт поступили Максим и Нина.* Но, конечно, не так: «*И Максим Нина поступили институт в пищевой*». Данная группа правил в синтаксисе носит название **правила порядка слов**.

Есть и другие синтаксические правила (например, правила соединения предложений в осмысленный текст). Однако первые две группы правил самые существенные: ведь, не построив предложение, не построишь и сам текст.

## **2. ПОНЯТИЯ СИНТАКСИЧЕСКОЙ ЕДИНИЦЫ И СИНТАКСИЧЕСКОЙ СВЯЗИ**

«Строительный материал» для синтаксиса (так сказать, «синтаксические кубики») мы будем называть синтаксическими единицами или синтаксическими объектами. Это слова, построенные из них словосочетания и также предложения, составленные из слов и словосочетаний (в особых случаях, например в неполных предложениях, в качестве синтаксических единиц рассматриваются определенного рода нулевые элементы).

Как нельзя построить дом без соединительных материалов, так нельзя из одних синтаксических единиц сконструировать предложение. Соединительным материалом в синтаксисе служат синтаксические связи (или отношения). Часть из них определяет главенствование одних единиц над другими и носит название **подчинительных связей**. Например, в сочетании *добрый вечер* слово *вечер* — главное, так как форма слова *добрый* (именительный падеж, единственное число, мужской род) обусловлена свойствами слова *вечер*. Другие устанавливают одинаковую синтаксическую роль элементов в составе целого. Это **сочинительные связи**. В сочетании *Петя и Катя* синтаксические функции слов *Петя* и *Катя* одинаковы.

Множество синтаксических связей предложения называется **синтаксической структурой** данного предложения.

## **3. СИНТАКСИЧЕСКИЕ СТРУКТУРЫ И ИХ ИЗОБРАЖЕНИЯ**

Конструкторам хорошо известно, что построение сложного объекта удобно начинать, когда под руками есть чертеж, т. е. модель будущего объекта. Синтаксисты также любят иметь дело с чертежами, а именно с графическим изображением синтаксической структуры. Здесь опять лингвистика прибегает к помощи математики. Если не знать некоторых важных математических понятий и не уметь ими пользоваться,

удобного синтаксического чертежа нарисовать не удастся. Разумеется, можно было бы обойтись и без чертежа, но оказывается (и это было замечательным открытием формальной лингвистики), что чертеж в синтаксисе не только отвечает целям наглядности, но и способствует обнаружению новых, ранее неизвестных языковых фактов.

### **3. А. Синтаксический граф**

Здесь на помощь лингвистам пришла относительно новая область математики — теория графов.

**Г р а ф о м** (точнее, **м у л ь т и г р а ф о м**) называется произвольное множество объектов с заданными на нем отношениями. Элементы этого множества называются **у з л а м и** или **в е р ш и н а м и** графа. Если узлами графа являются синтаксические единицы, а отношения являются синтаксическими, то граф называется **с и н т а к с и ч е с к и м**.

Граф обычно рисуют на плоскости. При изображении графа узлы рисуются в виде точек со стоящими при них надписями, а отношения изображаются в виде **с т р е л о к**, или, иногда говорят, **д у г**, направленных от первого члена отношения ко второму. Изображение графа тоже называют графом.

### **3. В. Синтаксическое дерево**

Несмотря на сравнительно простое определение, синтаксический граф реального предложения — это настолько сложный объект, что с ним трудно иметь дело при решении прикладных задач и, в частности, при построении синтаксических алгоритмов. В формальной лингвистике обычно имеют дело с относительно простым частным случаем графов — с деревьями. **Д е р е в о м** в теории графов называется граф, на который накладываются дополнительные ограничения: 1) существует единственный узел, в который не «входит» ни одна стрелка (этот узел называется **к о р н е м** дерева); 2) в каждый узел, кроме корня, «входит» ровно одна стрелка; 3) длинный граф не содержит «циклов» (т. е. невозможно, двигаясь из какого-нибудь узла в направлении стрелок, вернуться в тот же самый узел). Синтаксическое дерево отражает не все синтаксические связи, а только некоторые, в известном смысле наиболее важные.

### **4. ТИПЫ СИНТАКСИЧЕСКИХ ДЕРЕВЬЕВ**

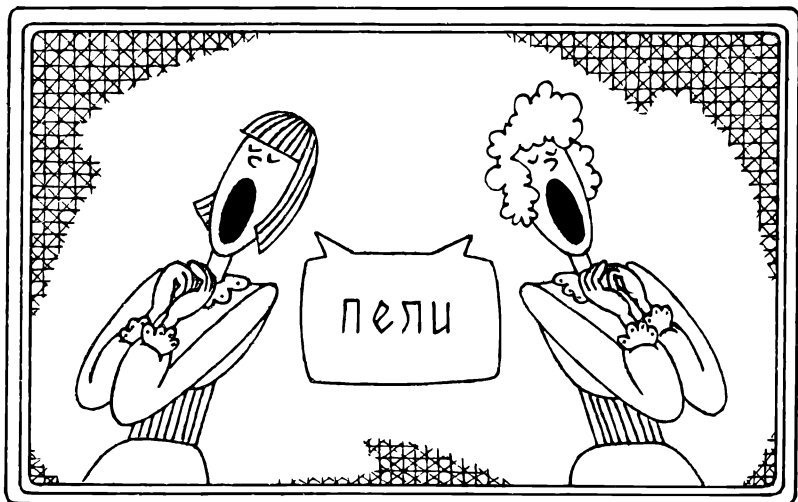
В зависимости от того, какие типы связей лингвисты считают нужным отразить в дереве, выделяются **дерево подчинения** и **дерево составляющих**.

#### 4. А. Дерево подчинения

**Формальное и смысловое подчинение.** Синтаксическое подчинение бывает двух видов: формальное и смысловое.

**Ф о р м а л ь н о е** подчинение состоит в том, что одна синтаксическая единица задает грамматическую форму слов в составе другой синтаксической единицы. Например, имя существительное определяет род, число, падеж согласованного с ним определения: *большой дом, большая река, большое поле, большие леса*. Глагол определяет выбор предлога вместе с падежной формой управляемого существительного: *дойти до сути, войти в дом, отойти от стола*.

**С м ы с л о в о е** подчинение состоит в том, что синтаксическая единица, обозначающая некоторую ситуацию, такую, что последняя характеризуется определенным набором «участников», подчиняет себе слова, обозначающие этих участников или сопутствующие обстоятельства. Так, в предложении *Девочки пели романс* слово *пели* обозначает ситуацию с двумя участниками. Это субъект действия, отвечающий на вопрос *кто пел?*, и объект действия, отвечающий на вопрос *что пел?* Тем самым глагол *пели* по смыслу подчиняет себе, соответственно, синтаксические единицы *девочки* и *романс*. В предложении *Она резала обои ножницами* глагол обозначает ситуацию, в которой три участника. Это субъект (*кто резал?*), объект (*что резал?*) и инструмент (*чем резал?*), а в предложении *Коля спит* слово *спит* обозначает ситуацию всего с одним участником-объектом.





Оба вида отношения подчинения, устанавливаемые между синтаксическими единицами, считаются синтаксическими **связями**. Направление этих связей определяется рядом заранее принимаемых соглашений. Отметим, что формальное и смысловое подчинение обычно имеют одинаковое направление, но иногда они направлены противоположно друг другу. Например, в приведенном выше предложении *Девочки пели романс* сказуемое *пели* семантически подчиняет *девочки*, а формально число слова *пели* определяется числом подлежащего *девочки*. Синтаксическое дерево, отражающее синтаксическое подчинение одних синтаксических единиц другим, называется **д е р е в о м п о д ч и н е н и я**. Таким образом, в узлах дерева подчинения стоят синтаксические единицы, а стрелки соответствуют разным видам синтаксического подчинения.

**Линейное упорядочение узлов дерева.** Оказывается, что математическое свойство «древесности» трудно совместить с реальным синтаксическим строением предложения, или, как говорят лингвисты, с его синтаксической структурой, в том смысле, что дерево подчинения не отражает всех возможных синтаксических связей в предложении. Одна из причин этого состоит как раз в том, что формальное подчинение часто не согласовано со смысловым. Приходится, если ставить перед собой цель сохранить «древесность», идти на некоторый компромисс. Сохранить дерево, вообще говоря, желательно, так как математическая и алгоритмическая обработка графа в этом случае упрощается. Каждый раз, оценивая степень важности конкретной синтаксической связи, лингвист принимает решение, отражать или не отражать ее в синтаксическом дереве.

Чертеж объекта считается удовлетворительным, только если по нему объект строится однозначно. Попробуем с этой точки зрения оценить синтаксические деревья. Восстанавливаются ли по ним предложения, для которых эти деревья были нарисованы?

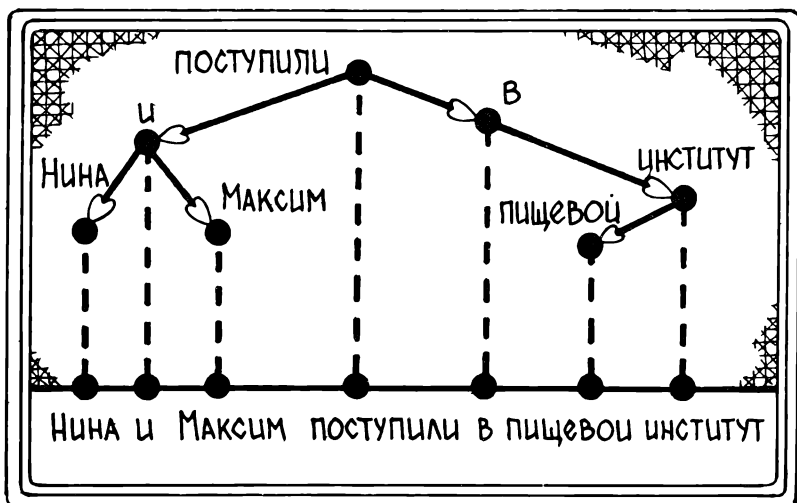
Понятно, что нет. Ведь предложение характеризуется не только синтаксическими связями, но и порядком слов. Одна и та же синтаксическая структура может давать предложения с разным порядком слов. Ср.: *Нина и Максим поступили в пищевой институт* и *В пищевой институт поступили Нина и Максим*.

Если мы хотим, чтобы по синтаксическому дереву однозначно прочитывалось предложение, нужны дополнительные формальные средства. Нам нужно каким-то образом линейно упорядочить узлы дерева. Если бы все узлы дерева всегда можно было изобразить в виде точек, лежащих на одной

прямой, то это было бы сделать просто: ведь точки на прямой уже линейно упорядочены. Но в общем случае изображение дерева имеет достаточно сложную конфигурацию, которую при расположении узлов на одной прямой трудно было бы воспринимать зрительно, особенно когда синтаксические связи изображаются в виде отрезков.

**Расположенное дерево.** Поэтому для упорядочения узлов дерева используют особую прямую, лежащую вне дерева и называемую *направляющей прямой* дерева, а также операцию *проекции* узлов дерева на эту прямую. А именно синтаксическое дерево подчинения называют *расположенным* относительно данной направляющей прямой, если при ортогональной проекции на эту прямую никакие два узла дерева не проецируются в одну точку. При такой проекции мы фактически используем свойство линейной упорядоченности точек прямой, а именно из двух узлов дерева левее будет тот, у которого проекция левее (см. рис.). (Легко доказать, что отношение «быть левее» на множестве узлов дерева при такой конструкции является транзитивным.)

Для любого дерева подчинения найдется такая прямая, относительно которой это дерево будет расположенным. Поэтому можно говорить о *расположенном* дереве подчинения, подразумевая при этом пару — дерево с фиксированной конфигурацией и некоторую прямую из множества всех прямых, относительно которых оно расположено. Содержательно направляющая прямая соответствует линейному порядку слов в предложении.



**Свойство проективности.** Если дерево подчинения расположено, то предложение восстанавливается по дереву однозначно, а именно в соответствии с порядком его узлов, причем синтаксические связи слов в предложении будут такими же, как и в дереве.

Соотношение между синтаксическими связями расположенного дерева подчинения и порядком слов в соответствующем предложении называется **свойством проективности**. Обнаружение этого свойства стало возможным лишь благодаря взаимодействию лингвистики с математикой. Действительно, без введения формальных понятий дерева и расположенного дерева свойство проективности строго сформулировать не удастся.

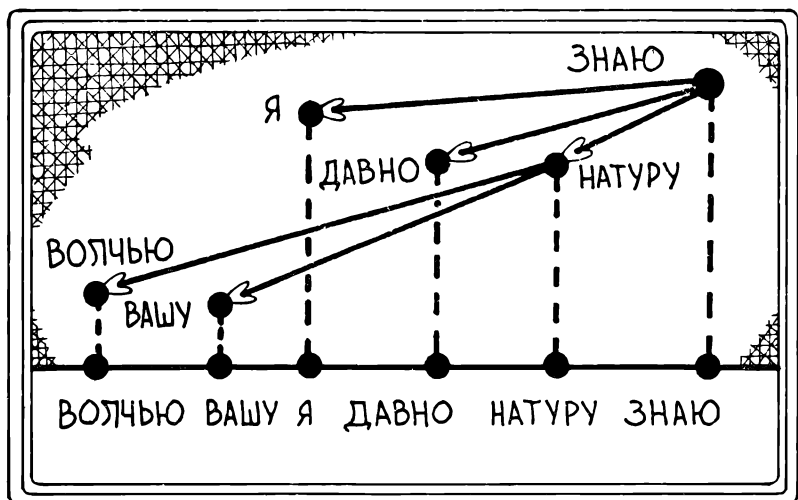
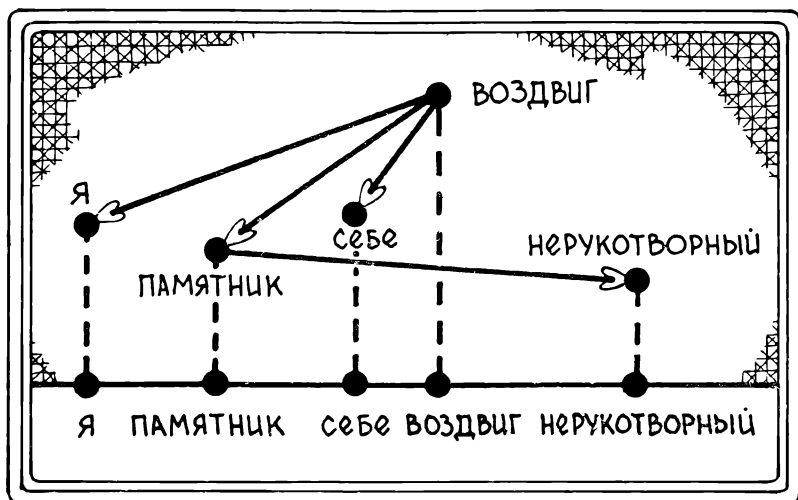
Среди всех допустимых порядков слов в предложении одни расположения ощущаются как вполне нормальные, а другие вызывают неловкость и даже неприятие. Попытки лингвистов объяснить, в каких случаях и почему возникают подобные ощущения, и привели к установлению связи между порядком слов и синтаксическими отношениями.

**Проективно расположенное дерево.** Как же устанавливается такая связь? Чтобы ответить на этот вопрос, нужно сначала условиться, как мы будем изображать синтаксические отношения. Мы уже видели, что удобно их изображать в виде стрелок, идущих от первого члена отношения ко второму. Теперь мы можем дать формальное лингвистическое определение свойства проективности.

Расположение дерева подчинения называется **проективным**, если при указанном соглашении перпендикуляры, опущенные из узлов на направляющую прямую, не «задевают» дерева, т.е. ни один из этих перпендикуляров не пересекает ни одной стрелки.

Оказывается, что при таком понимании проективности предложения с допустимым порядком слов в подавляющем большинстве случаев имеют проективные деревья, а непроективность дерева отвечает не вполне удачному расположению слов во фразе. Для проективных деревьев синтаксически связанные слова расположены, как правило, максимально близко друг от друга и не отделяются мешающими понять их связью словами.

**Иллюстрация свойства проективности.** Предложение *Данная книга предназначена для ребят, интересующихся лингвистикой* имеет проективное дерево подчинения. Рассмотрим, что же произойдет при некоторых перестановках слов в этом предложении. Предложение *Данная книга предназначена для интересующихся лингвистикой ребят* имеет также проективную структуру, а неудачное предложение *«Данная книга*



предназначена для интересующихся ребят лингвистикой» непроективную (проверьте это!).

Не следует думать, что проективность сама по себе всегда обеспечивает правильный порядок слов. Располагая слова во фразе, следует учитывать не только проективность, но и общие законы языка (например, в современном литературном языке нельзя ставить предлог *для* после существительного, к которому он относится), место данного предложения в тексте, значимость отдельных элементов (*свет и тьма, добро и зло*

лучше, чем *тьма и свет, зло и добро*). Однако существенно, что если все перечисленные закономерности приняты во внимание, то проективность дерева подчинения является характеристическим свойством правильного расположения слов в предложении.

Правда, есть область, где проективность часто нарушается, причем вполне сознательно. Речь идет о поэзии, свободной от многих норм обычного языка. Порядок слов здесь подчинен особой ритмике стиха, требованиям стихотворного размера, заостренному вниманию на том или ином образе или художественном приеме. Благодаря непроективным структурам (см. рисунок на с.113) достигается возвышенность, приподнятость стиля (*Я памятник себе воздвиг нерукотворный*) или, наоборот, имитируется «сниженная», разговорная речь (*Волчью вашу я давно натуру знаю*).



44. Исправьте следующие предложения, поменяв порядок слов так, чтобы у полученных предложений дерева подчинения были проективными: *Очень коня любил один шах своего; Я тайной владею обучения человеческому животных языку.*

45. Приведенные ниже предложения взяты из разных поэтических произведений. Покажите, что их деревья непроективны. Для каждого предложения найдите такую перестановку слов, чтобы результирующее предложение имело проективную структуру.

1) Ноги я Твои в подол уперла. (Пастернак.) 2) Кожей он твоей покроет мне вспотевшие бока. (Пушкин.) 3) Легкокрылая мне изменила радость. (Пушкин.) 4) ...Над отечеством свободы просвещенной взойдет ли наконец прекрасная заря? (Пушкин.)

#### 4. В. Дерево составляющих

**Понятие дерева составляющих.** При помощи дерева подчинения, как мы видели, фиксируются связи между словами. Но ведь существуют синтаксические единицы, более крупные, чем слова, и между ними тоже логично предположить наличие каких-то связей. Непосредственно отразить эти связи призвано **д е р е в о с о с т а в л я ю щ и х**.

Рассмотрим предложение *Он любит только жареные пирожки*. Можно ли считать, что слово *только* связано лишь с каким-то одним словом? При одном понимании (с интонационно выделенным словом *жареные*) можно считать, что да: *только* связано с *жареные* (ср.: *Он любит не только жареные, но и печеные пирожки*). Но при другом понимании разумно предполагать, что *только* связано с целой группой слов — *жареные пирожки* (ср.: *Он любит не только жареные пирожки, но и кислую капусту*). Для непосредственного отражения этой последней связи дерева подчинения недостаточно. Нужны либо средства, дополнительные к дереву подчинения и не

входящие в него, либо новый объект, элементами которого будут синтаксические единицы, в общем случае большие, чем слово. Лингвисты построили такой объект и назвали его деревом составляющих.

Мы не будем давать определения этого понятия, а изложим его общую идею.

Суть ее заключается в последовательном вычленении более мелких частей из более крупных.

Когда мы отвечаем на вопрос, из чего состоит данный объект, то обычно не стремимся сразу назвать его самые мелкие части. Так, мы не говорим, что стул состоит из элементарных частиц, а указываем в качестве его составных частей (составляющих) достаточно крупные: спинку, сиденье, ножки. До мельчайших частей мы доходим, лишь последовательно отвечая на серию аналогичных вопросов, каждый раз задаваемых применительно к уже выделенным частям.

Так, минеральная вода состоит из воды и растворенных в ней минеральных солей. Вода, в свою очередь, представляет собою собрание отдельных молекул; каждая молекула состоит из двух атомов водорода и одного атома кислорода. Атом состоит из ядра и электронов и т. д.

Эту естественную идею последовательной членимости лингвисты перенесли на языковые объекты. Вспомним, например, разбор слова по составу. Сначала выделяют основу и окончание, затем, если основа далее членима, выделяют в ней базовую основу и словообразовательные морфемы, далее членят базовую основу и т. д. (см. четвертую главу).

Такой же метод лежит в основе построения синтаксических деревьев составляющих. Здесь под составляющими понимают синтаксические составные части предложения.

Рассмотрим предложение *Важные лингвистические открытия нередко незаслуженно отвергались*. Из каких частей оно состоит? Лингвисты начинают деление с того, что выделяют так называемые группу подлежащего и группу сказуемого. Группу подлежащего здесь составляет часть *важные лингвистические открытия*, а группу сказуемого — *нередко незаслуженно отвергались*. Группа подлежащего далее делится на составляющие *важные* и *лингвистические открытия*, а *лингвистические открытия* — на *лингвистические* и *открытия*. Ясно, что на этом деление группы подлежащего заканчивается. Самые мелкие составляющие носят название атомарных или терминальных. До терминальных составляющих в группе сказуемого можно дойти, выделив в ней составляющие *нередко* и *незаслуженно отвергались*, а в последней — составляющие *незаслуженно* и *отвергались*. Процесс и результаты деления наглядно представляются в виде дерева, которое и



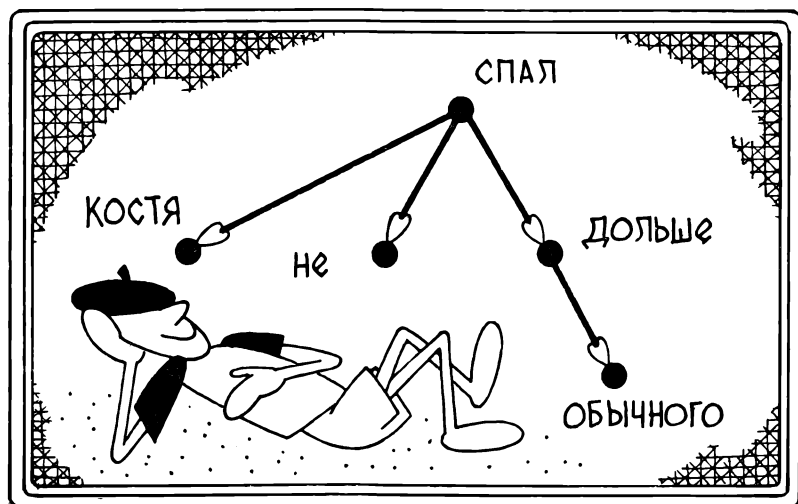
В отличие от дерева подчинения дерево составляющих с помощью особого приема — скобочной записи дерева — можно изобразить еще более экономным с графической точки зрения способом. Например: ((Важные (лингвистические открытия)) (нередко (незаслуженно отвергались))).

Отметим один важный момент. Скобочную запись дерева составляющих мы можем построить только для такого дерева, у которого составляющие не «разрываются» (утверждение, что составляющие не «разрываются», означает, что для любых двух слов, входящих в какую-либо одну составляющую, все слова между ними входят в ту же составляющую). Доказательство этого факта достаточно сложно и поэтому здесь не приводится. В данной ниже задаче имеется в виду построение алгоритмов для деревьев без «разрывных» составляющих.

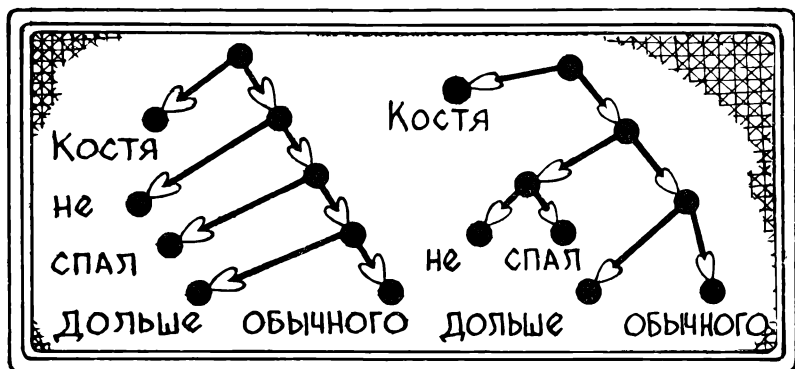


46. Построить два алгоритма: алгоритм перехода от древесного изображения структуры составляющих к скобочной записи дерева и алгоритм обратного перехода.

**Отражение смысловых различий в дереве составляющих.** С помощью дерева составляющих иногда удается фиксировать разницу по смыслу у некоторых синтаксических единиц. Рассмотрим предложение *Костя не спал дольше обычного*. Оно может быть понято двумя способами: 1) Костя не находится в состоянии сна дольше, чем обычно (не было так, что Костя спал дольше обычного, т. е. Костя проснулся не позже, чем просыпался обычно); 2) Костя бодрствовал дольше, чем







обычно (было так, что Костя обходился без сна дольше обычного, т. е. Костя заснул позже, чем обычно). Этим двум пониманиям соответствует одно и то же синтаксическое дерево подчинения (см. рисунок на с.117), но разные деревья составляющих (см. рисунок на с.118).

При первом понимании этого предложения составляющая *не спал дольше обычного* членится на *не* и *спал дольше обычного*, т.е. отрицание воздействует на всю составляющую *спал дольше обычного*. При втором понимании членение иное: составляющая *не спал* соединится с составляющей *дольше обычного*, а составляющей *спал дольше обычного* вообще нет. Поскольку *не спал* означает 'находился в состоянии бодрствования', то мы и получаем второе понимание.

Как мы видим, дерево составляющих иногда может помочь в тех случаях, когда построение дерева подчинения не приводит к желаемым результатам.



47. Рассмотрим предложение *Магазин принимает от населения старые журналы и книги*.

Это предложение можно понять двумя способами, каждому из которых отвечает своя структура составляющих. Нарисуйте два дерева составляющих и укажите те смыслы, которые отражены при помощи этих деревьев.

## II. ПРИНЦИП СИНТАКСИЧЕСКОЙ ПРОСТОТЫ

### 1. ПРОСТЫЕ И СЛОЖНЫЕ ФРАЗЫ

#### 1. А. Синтаксическая громоздкость фразы

С интуитивной точки зрения, решение вопроса о том, проста фраза или тяжеловесна, не составляет труда. Мы обычно не останавливаемся в размышлении, читая в книгах высказывания типа «Чехов пишет легко и просто, а у Толстого

манера письма тяжеловесна». Однако лингвисту, занимающемуся формальным анализом естественного языка, такого рода расплывчатые утверждения равным счетом ничего не объясняют. Поэтому он стремится придать любым суждениям математическую четкость и строгость. Один из способов достичь этого состоит в том, чтобы построить математическую теорию синтаксической громоздкости фразы. В рамках этой теории точный смысл приобретают такие вопросы, как: «Что такое громоздкая фраза?», «Какие бывают виды громоздкости?», «Какая из двух фраз является более (менее) громоздкой?»

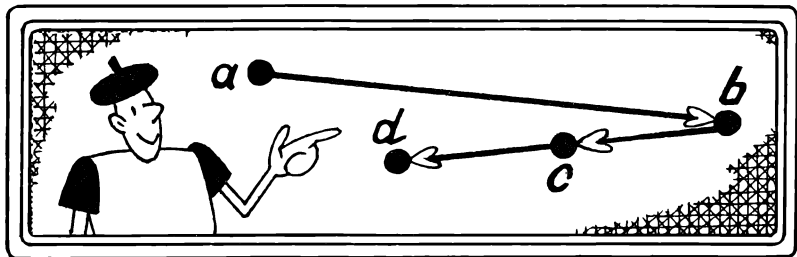
Основным инструментом теории синтаксической громоздкости являются деревья. Надеемся, что читателю, ознакомившемуся с содержанием предыдущей главы, это покажется вполне естественным. Действительно, структурная простота фразы определяется источником и характером связей между синтаксическими единицами, а дерево как раз и было придумано для представления этих связей.

### 1. В. Причины тяжеловесности фразы

Одна из причин тяжеловесности фразы лежит в неестественном порядке слов — то ли неудачном, то ли преднамеренно изощренном. Вторая причина — наличие чересчур большого числа дополнительных связей, нарушающих основной ход мысли автора.

Пример неудачного порядка слов дает предложение *Здесь изображено отражающее устанавливаемое между входящими в предложение словами дерево*. Изменив порядок слов, получим менее громоздкое и более приятное предложение *Здесь изображено дерево, отражающее связи, устанавливаемые между словами, входящими в предложение*.

Примером предложения с большим числом дополнительных связей может служить предложение *Расположенный недалеко от центра дом, изобилующий архитектурными украшениями, построенный еще в прошлом веке, был очень красивым*. Здесь, как ни меняй порядок слов, громоздкость не устроишь. Приходится прибегать к более радикальным средст-



вам — например, к разбиению фразы на две или несколько более мелких.

Эти общие характеристики громоздкости получают наглядную интерпретацию в дереве подчинения.

Введем ряд вспомогательных определений.

Говорят, что узел *b* дерева подчинения *расположен между* узлами *a* и *c* того же дерева, если *b* левее одного из них, но правее другого. Например, на рисунке на с.119 узел *c* расположен между узлами *a* и *b*, а также между узлами *d* и *b*; узел *d* расположен между узлами *a* и *b*, а также между *a* и *c*. Длинной покрытия стрелки назовем число узлов между ее началом и концом. На рисунке длина стрелки *ab* равна 2, а длина покрытия стрелки *bc*, так же как длина покрытия стрелки *cd*, равна 0 (эти стрелки не «покрывают» ни одного узла).

Большая длина покрытия стрелки означает большую удаленность подчиняемого слова от подчиняющего. Большая удаленность связанных слов затрудняет восприятие фразы и утяжеляет ее.

Будем говорить, что одна стрелка *обрамляет* другую, если начало и конец второй стрелки лежат между началом и концом первой. В частности, на рисунке стрелка *ab* обрамляет стрелки *bc* и *cd*, а сами стрелки *bc* и *cd* не обрамляют никакой стрелки.



48. Доказать, что отношение обрамления на множестве стрелок транзитивно.

Основное утверждение синтаксической теории громоздкости состоит в том, что, чем больше обрамлений в дереве подчинения, тем более тяжеловесной выглядит фраза.

Становится понятным, чего должен избегать автор, который стремится сделать свой текст более простым. Он должен уменьшать суммарную длину покрытий и количество обрамлений.

## 2. ГИПОТЕЗА ИНГВЕ

### 2. А. Объем оперативной памяти

Исследование сложности дерева привлекло к себе внимание не только лингвистов, но и психологов. Это было прежде всего связано с тем, что громоздкая фраза трудна для понимания и запоминания; по ней нелегко следить за мыслью автора.

В психологии давно известен факт, что объем оперативной, т. е. непосредственной, памяти человека ограничен числом  $7 \pm 2$ , т. е. человек одновременно может оперировать не более чем с пятью—девятью символами. Если средний человек слышит последовательность слов, чисел или каких-то

других знаков и не видит никакой связи между ними, то он может запомнить максимум девять таких символов.

## 2. В. Оперативная память и синтаксическая структура предложения

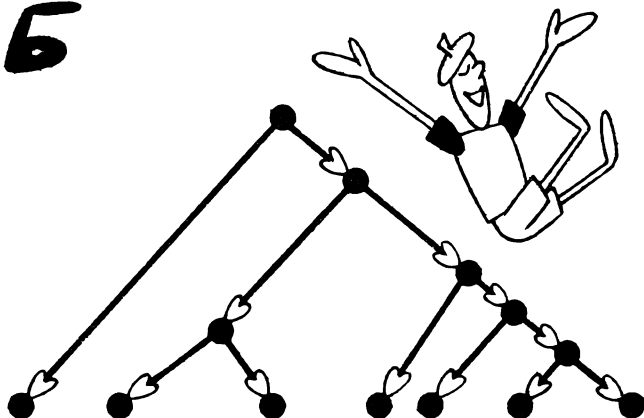
Американский ученый В.Ингве попытался соотнести ограниченность человеческой памяти, проявляющуюся в процессе порождения или восприятия речи, с некоторыми характеристиками дерева составляющих. В основе гипотезы Ингве, суть которой мы попытаемся изложить ниже, лежат следующие допущения:

1. Единицами производства речи и ее понимания являются составляющие предложения; процесс построения (понимания) предложения есть в большой степени процесс построения (анализа) структуры его составляющих.

**А**

( ПЕТР (( БЫСТРО ПРИБЛИЖАЛСЯ)  
( к ( МЕСТУ ( СВОЕГО НАЗНАЧЕНИЯ) ) ) ) ) )

**Б**



ПЕТР БЫСТРО ПРИБЛИЖАЛСЯ К МЕСТУ СВОЕГО НАЗНАЧЕНИЯ

2. Развертывание предложения во времени обуславливает необходимость хранить в непосредственной памяти некоторые символы, соответствующие предложению и его составляющим.

3. Пусть синтаксическая единица *A* разбивается на составляющие  $\alpha, \beta, \gamma, \dots$  ( $\alpha, \beta, \gamma, \dots$  в свою очередь, могут разбиваться на составляющие). Тогда  $\alpha, \beta, \gamma, \dots$  называются непосредственными составляющими единицы *A*. Допущение состоит в том, что объектами запоминания являются непосредственные составляющие.

При развертывании синтаксической единицы (в частности, предложения) во времени до построения (анализа) последней из ее непосредственных составляющих мы должны хранить в оперативной памяти все предыдущие. В противном случае структура синтаксической единицы останется неопределенной.

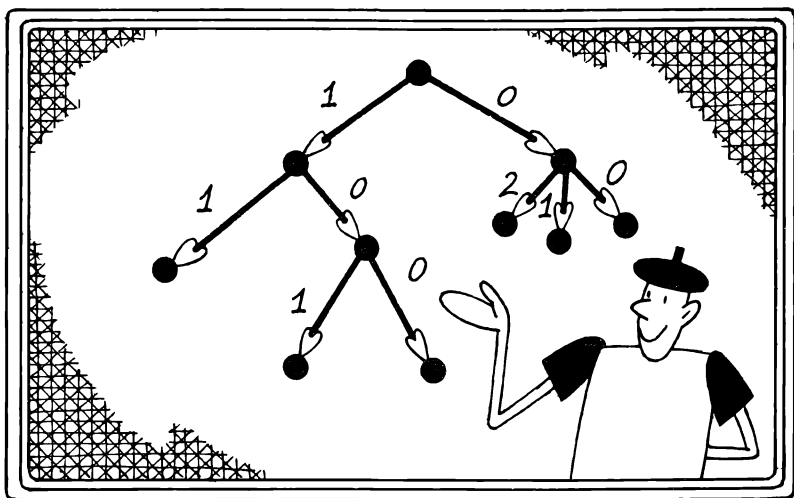
Рассмотрим один пример. Предложение *Петр быстро приближался к месту своего назначения* имеет следующую структуру (см. рисунок на с.121). Здесь (А) соответствует скобочной записи, а (Б) — изображению в виде дерева.

Возьмем синтаксическую единицу — *быстро приближался к месту назначения*. Ее непосредственные составляющие — *быстро приближался* и *к месту назначения*. Очевидно, что, пока мы говорим (или слышим) *быстро приближался*, не говоря или не понимая куда, нам приходится эту составляющую держать в оперативной памяти. Как только мы начнем говорить, куда приближался Петр, мы можем отправить составляющую *быстро приближался* в долговременную память: ведь человек не может перегружать оперативную память, он должен оставить в ней место для новой поступающей информации.

Уже из этих допущений читатель, наверно, догадался о наличии связи между памятью человека и структурой предложения. Утверждение о наличии такой связи и составляет содержательную сторону гипотезы Ингве. К сожалению, при всем своем изяществе гипотеза Ингве в общем случае неверна. Она отвергается фактами целого ряда языков (венгерского, латышского и др.). Тем не менее мы включили раздел, рассказывающий о ней, в книгу, поскольку значение исследований Ингве и его последователей для развития лингвистики (в частности, синтаксических учений) и психологии трудно переоценить.

Между тем сам Ингве не ограничился одной лишь констатацией связи, а дал ей количественную оценку. Он выявил общий объем оперативной памяти, необходимый для построения (понимания) конкретного предложения.

Для этого Ингве предложил пронумеровать все ветви



(стрелки) дерева по следующему правилу: а) ветви, выходящие из каждого узла, нумеруются справа налево числами 0, 1, 2, 3, ...; б) переходя от одного узла к другому, нумерацию ветвей начинают заново (см. рисунок).

В.Ингве строил свою процедуру для расположенных деревьев составляющих, т.е. деревьев, у которых все терминальные узлы расположены в дереве в соответствии с их линейным порядком: из двух терминальных узлов левее тот, помета которого (слово) идет в предложении раньше. Такие деревья могут, вообще говоря, содержать и «разрывные» составляющие, но для таких деревьев гипотеза Ингве проходит лишь с известным приближением.

Номер при каждой стрелке указывает, сколько стрелок, выходящих из того же узла, имеется справа от нее. Иными словами, номер говорит, сколько ячеек оперативной памяти требуется зарезервировать для построения (или анализа) синтаксической единицы, соответствующей данному узлу.

### 3. ГЛУБИНА ДЕРЕВА, ПО ИНГВЕ

Введем формальное понятие **г л у б и н ы д е р е в а**.

Предположим, что все ветви дерева составляющих пронумерованы по правилам, приведенным выше. Выберем произвольный узел и подсчитаем сумму чисел, написанных на стрелках, ведущих к нему от корня. Эта сумма называется **г л у б и н о й д а н н о г о у з л а**. Максимальная из глубин всех узлов дерева называется **глубиной дерева**.

При выработанном способе нумерации **большая глубина**

означает сильное ветвление дерева влево, а меньшая — ветвление вправо.

Лингвисты на опыте обнаружили, что хорошим, легко запоминающимся фразам независимо от их длины соответствует малая глубина дерева. Чаще всего глубина равна 2 или 3 и практически никогда не превышает 8.

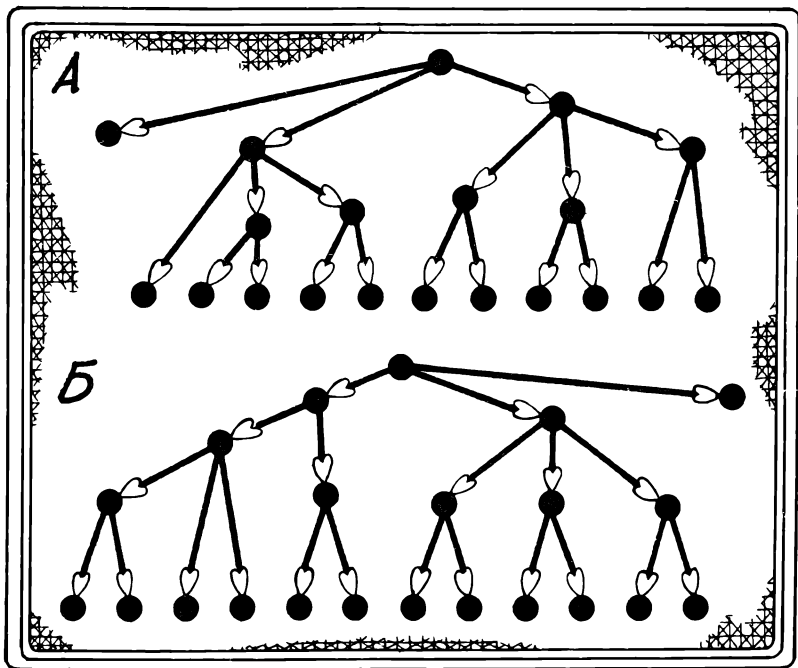
Чему же содержательно отвечает глубина дерева? Глубина узла — это максимальное число единиц запоминания, которые надо удержать в оперативной памяти до тех пор, пока данный узел не отправлен в долговременную память. Тем самым глубина дерева интерпретируется как объем оперативной памяти, необходимый для запоминания всего предложения. Оказалось, что этот объем не превосходит числа 9.

Мы видим, что существует замечательная связь между лингвистическими фактами, психологическими наблюдениями и математическими конструкциями. Очевидно, что обнаружение этой связи есть результат взаимодействия трех дисциплин: лингвистики, математики и психологии.



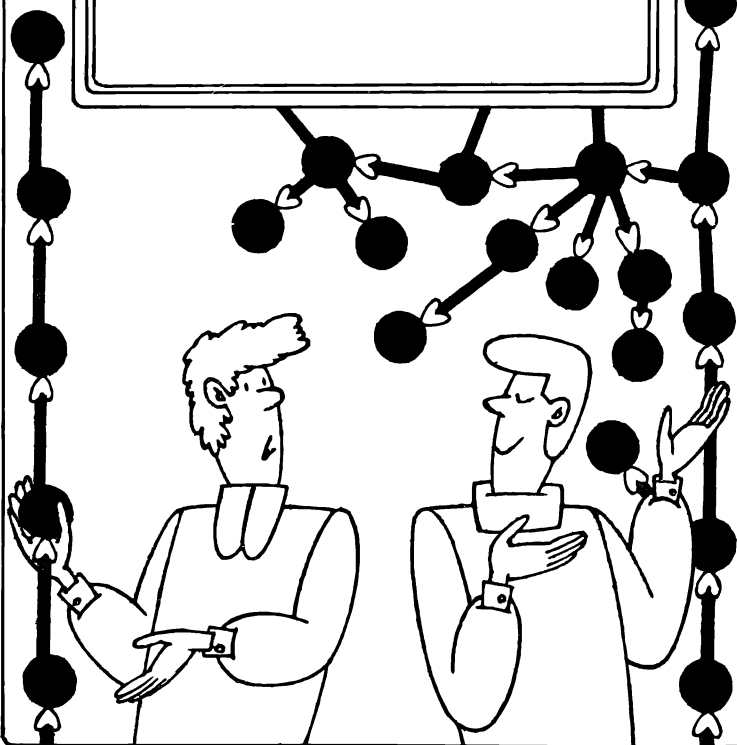
49. 1) Найти глубину каждого из деревьев на рисунке.

2) Построить алгоритм, который для произвольного дерева будет вычислять его глубину.



Глава седьмая

**Математика  
и анализ  
художественного  
текста**





# 1. СТИЛИСТИЧЕСКАЯ ДИАГНОСТИКА

## 1. ИНДИВИДУАЛЬНЫЙ СИНТАКСИС ПИСАТЕЛЯ

**Е**сли бы, скажем, в XIX или даже в начале XX в. кто-нибудь поставил вопрос, можно ли использовать математические методы при анализе художественного текста, то этот вопрос вызвал бы недоумение даже у самых образованных людей. Ныне же, благодаря обнаружению формальных приемов анализа текста, ситуация в корне изменилась. Строгие и точные методы изучения художественных текстов позволили найти в них свойства, которые было трудно заметить «невооруженным глазом». Ниже мы покажем, что рисунок дерева подчинения является формальной характеристикой индивидуальной манеры письма. Здесь синтаксис смыкается с другой областью исследования текста — со стилистикой. При этом литературоведа, занимающегося проблемами стиля, интересуют прежде всего такие содержательные параметры, как сюжет, идейная направленность, характеры и поступки героев, основная мысль произведения, тема и т. д.; а внимание лингвистов обращено на языковые средства выражения содержательных характеристик, на «словесную ткань» произведения.

У искушенного читателя есть интуитивное представление о творческом почерке писателя. Такой читатель знает его излюбленные приемы и среди них типичные для данного писателя способы построения фраз, ритмику, акценты и т. д., т. е. индивидуальный синтаксис автора. Знание основных черт индивидуального синтаксиса может понадобиться и при решении обратной задачи — по данному тексту определить, кто его автор. Эту задачу называют задачей атрибуции текста. Ее приходится решать историкам, литераторам, которым в руки попадает текст неизвестного автора, работникам литературных музеев, библиотекарям.

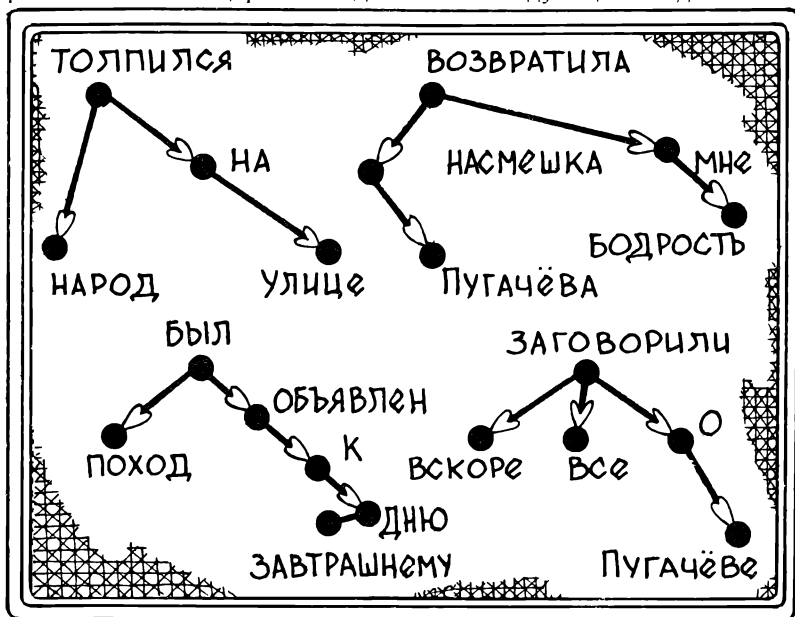
Основной гипотезой при формальном решении этой задачи служит предположение, что индивидуальный синтаксический почерк отражается в изображении деревьев для структур наиболее типичных для автора фраз. Использование ЭВМ позволило лингвистам, опираясь на вероятностные и статистические методы анализа, по виду деревьев и их количественным характеристикам устанавливать с большой достоверностью, кто автор данного текста.

## 2. ПЕРЕВОД ОСОБЕННОСТЕЙ АВТОРСКОГО СТИЛЯ НА ФОРМАЛЬНЫЙ ЯЗЫК СИНТАКСИЧЕСКИХ ДЕРЕВЬЕВ

Покажем на примере творчества нескольких писателей, как на язык деревьев переводятся трудноуловимые и на первый взгляд неформализуемые особенности стиля, которые кладутся в основу стилистической диагностики.

### 2. А. Пушкинская, лермонтовская и гоголевская фразы — сходства и различия

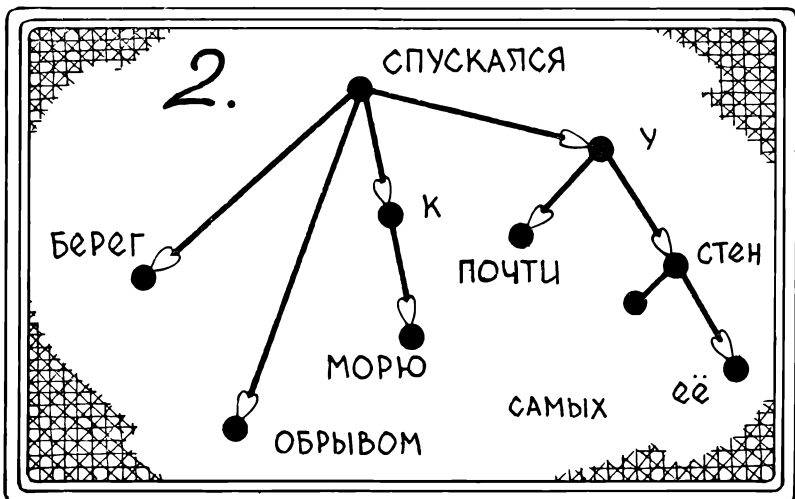
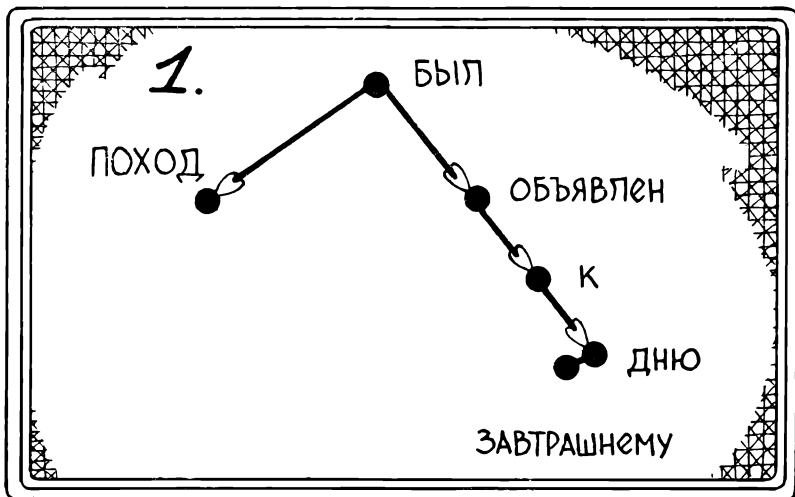
Основная черта синтаксиса прозы А.Пушкина — ее ритмизованность и подчиненный ей лаконизм выражений. В прозаических произведениях Пушкина преобладают краткие фразы, часто встречаются нераспространенные предложения. Так, если взять «Капитанскую дочку», то для нее типичны расположенные деревья подчинения следующего вида:

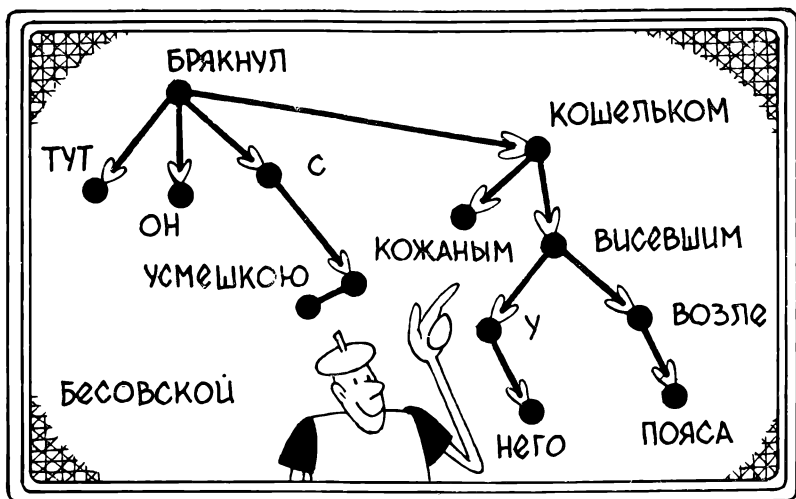


Пушкинский текст в основном состоит из предложений, в которых не более 11 слов, а рисунки их деревьев либо симметричны, либо имеют длинный правый отросток. При этом даже для длинных фраз громоздкие деревья практически не возникают. Как мы видим, интуитивное ощущение прозрачности пушкинской фразы соответствует строгому понятию синтаксической простоты.

Деревья лермонтовской прозаической фразы во многом похожи на пушкинские, хотя подсчеты показывают, что в

среднем предложения у Лермонтова чуть-чуть длиннее и чуть-чуть сложнее. Впрочем, есть важное различие в рисунках деревьев, свойственных этим авторам. Ширина ветвления у корня дерева (см. рисунок) для фразы из «Героя нашего времени» гораздо больше, чем для фразы из «Капитанской дочки». Это означает, что дерево лермонтовской фразы растетвширь, тогда как в пушкинской фразе оно растет вглубь. Большая ширина ветвления возникает вследствие того, что



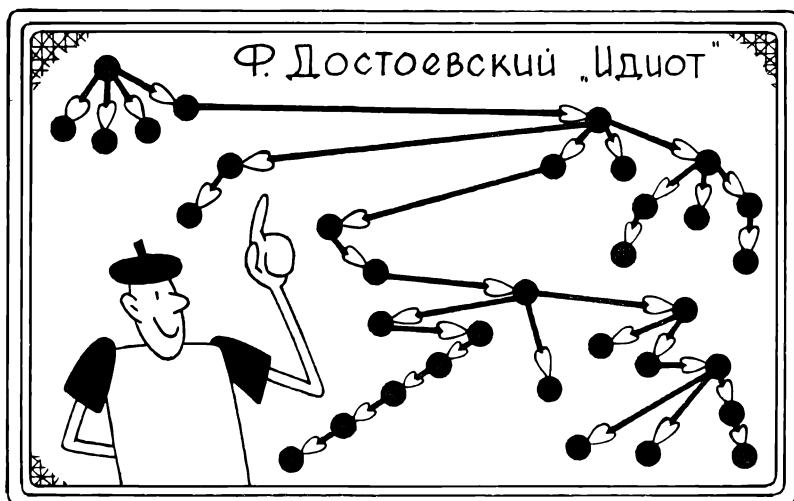
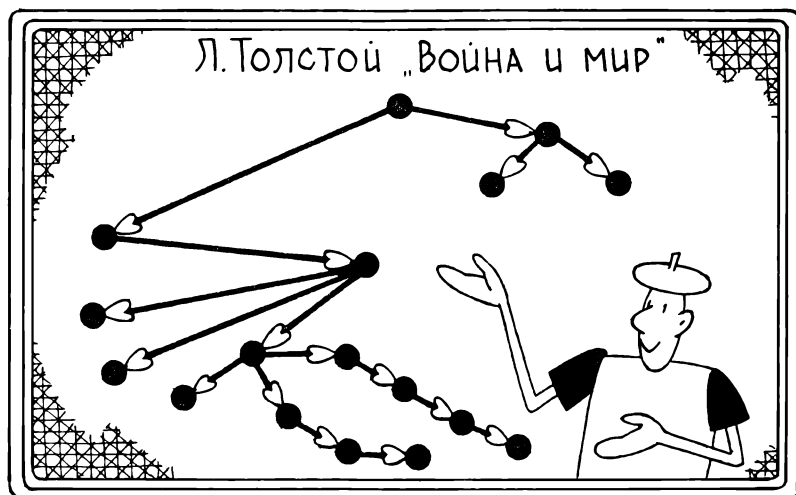


сказуемые в лермонтовской фразе подчиняют себе не только дополнения, но и разнообразные по структуре и значению обстоятельства. Из двух деревьев на рисунке первое более типично для Пушкина, а второе — для Лермонтова.

У Гоголя в «Вечерах на хуторе близ Диканьки» в стилистической пестроте фраз встречаются не только короткие предложения. Здесь в большом количестве представлены предложения, длинные и сложные по структуре. Даже относительно короткие предложения из 6—12 слов строятся у Гоголя весьма разнообразно: в его произведениях можно найти едва ли не любую теоретически возможную из данного числа слов конфигурацию ветвей дерева. В частности, наблюдается своеобразный тип структуры с многократными зигзагами, когда, спускаясь вниз по стрелкам, мы постоянно изменяем направление движения (см. рисунок).

## 2. В. Своеобразие синтаксиса Л.Толстого и Ф.Достоевского

Сравнив прозу Л.Толстого и Ф.Достоевского, можно отметить, что стилю обоих писателей присущи деревья достаточно сложной конфигурации. Но фразы Л.Толстого мы скорее опишем как громоздкие, а построения Достоевского как «неупорядоченные». Не случайно Ю.Нагибин как-то заметил, что язык Л.Толстого тяжеловесен, а язык Достоевского хаотичен. Как же переводится эта импрессионистическая оценка современного писателя на формальный язык деревьев? Громоздкость синтаксиса Л.Толстого сказывается в значительном ветвлении деревьев влево (что, как мы помним,



повышает глубину дерева — см. главу 6); «хаотичность» синтаксиса Достоевского, в свою очередь, проявляется в большом числе зигзагов в правой части дерева. На рисунке показан общий вид деревьев, характерных для этих писателей.

Мы надеемся, что уже этого немногочисленного иллюстративного материала достаточно, чтобы читатель понял, как лингвист определяет общий контур деревьев, отражающих индивидуальный стиль автора. Однако для вероятностно-синтаксического анализа все эти расплывчатые зрительные

представления требуется свести к определенному, по возможности небольшому, набору диагностических признаков. Известный лингвист И.П.Севбо, на результаты исследований которой мы опираемся в этой главе, предложила семь таких признаков. Приведем для примера пять из них.

1. Количество узлов дерева (т. е. количество слов во фразе).

2. Количество простых предложений в сложном. Для того чтобы учесть этот параметр, приходится особым образом пометить стрелки, соответствующие связям между частями сложного предложения.

3. Число уровней в дереве. Чтобы воспользоваться этим признаком, нужно предварительно определить понятия пути и длины пути. Последовательность узлов дерева  $a_1, a_2, \dots, a_n$  называется путем из  $a_1$  в  $a_n$ , если для любого  $i$  ( $1 \leq i \leq n-1$ ) из  $a_i$  в  $a_{i+1}$  идет стрелка, т. е. любые два соседних узла в этой последовательности соединены стрелкой. Число  $n$ , т. е. число узлов в пути, называется его длиной. Число уровней в дереве — это длина самого длинного из путей дерева.

4. Ширина ветвления корня, т. е. число узлов, подчиненных корню.

5. Максимальное количество перемен направления пути (зигзагов) в дереве. Учитываются два вида направления пути: левое и правое. Говорят, что путь на участке  $a_i, a_{i+1}$  направлен налево, если  $a_{i+1}$  расположен левее  $a_i$ . Аналогично определяется правое направление.

Мы видим, что на смену неформальному словесному описанию деревьев подчинения, свойственных тому или иному писателю, приходит синтаксическое исследование диагностических параметров. Вместо нечислового объекта — дерева — рассматривается кортеж из семи чисел, характеризующих строение соответствующего предложения. В этом случае индивидуальный синтаксис писателя может быть закодирован совокупностью числовых кортежей, взятых для наиболее типичных его предложений.

## II. УСТРОЙСТВО ТЕКСТА ВОЛШЕБНОЙ СКАЗКИ КАК РАЗНОВИДНОСТИ ТЕКСТА

### 1. ПОНЯТИЕ СЮЖЕТНОЙ ФОРМУЛЫ

Выше мы видели, что точные методы исследования текста возможны там, где есть большая повторяемость языковых единиц. Поэтому ученые — лингвисты и литературоведы — пытались найти для своих исследований такой материал, где повторяемость была бы выражена наиболее

ярко. В этом случае логическое сопоставление и количественный анализ в сочетании с собственно филологическими подходами и приводят к желаемому результату.

Удобным полигоном для испытания точных методов анализа литературных произведений послужили фольклорные тексты, и прежде всего народные волшебные сказки. Для многих из них характерно сходство тем, сюжетных линий, одни и те же герои. Персонажи разных сказок обладают сходными чертами характера, участвуют в сходных действиях, попадают в сходные ситуации. Последнему наблюдению ученые научились придавать точный смысл, моделируя интуитивное, неформальное понятие сходства с помощью вполне строгого понятия сюжетной формулы.

## **2. ОБЩАЯ ХАРАКТЕРИСТИКА СЮЖЕТНЫХ ФОРМУЛ РУССКОЙ ВОЛШЕБНОЙ СКАЗКИ**

### **2. А. Принципы построения сказки**

Впервые понятие сюжетной формулы ввел в научный обиход один из крупнейших фольклористов XX века В.Я.Пропп. Говоря о сюжетной формуле как о «функции» действующих лиц, Пропп сформулировал три основных принципа построения русской народной волшебной сказки: 1) эти сказки обладают постоянными, устойчивыми элементами, которые представляют собой функции действующих лиц независимо от того, кем и как они выполняются; 2) число таких функций ограничено; 3) порядок следования функций (сюжетных формул) всегда фиксирован.

### **2. В. Список сюжетных формул и комментарии к ним**

Главной научной заслугой Проппа явилось выделение сюжетных формул и описание условий и способов их реализации. Оказалось, что тридцати одной формулы достаточно, чтобы описать сюжет любой русской народной волшебной сказки. Чтобы читатель мог сам попытаться применить эти формулы к какой-нибудь из запомнившихся ему сказок, мы приведем их полный список с комментариями:

1) отлучка кого-либо из членов семьи; 2) запрет, обращенный к герою; 3) нарушение запрета; 4) выведывание; 5) выдача; 6) подвох; 7) невольное пособничество; 8) беда; 9) посредничество; 10) начинающееся противодействие; 11) герой покидает дом; 12) будущий даритель испытывает героя; 13) герой реагирует на действия будущего дарителя; 14) получение волшебного средства; 15) герой доставляется к месту нахождения предмета поисков; 16) герой и его противник вступают в борьбу; 17) героя метят; 18) противник побежден; 19) беда ликвидируется; 20) возвращение героя; 21)

героя преследуют; 22) герой спасается от преследования; 23) герой неузнанным прибывает домой или в другую страну; 24) некий ложный герой предъявляет необоснованные притязания; 25) герою предлагается трудная задача; 26) задача решается; 27) героя узнают; 28) ложный герой или противник изобличается; 29) герою дается новый облик; 30) ложный герой или противник наказывается; 31) герой вступает в брак.

1. В формулах названы — через их сюжетные роли — основные действующие лица сказки. Это герой, даритель, противник и ложный герой. Кроме того, в формулах участвуют лица, прямо не названные. Это, например, любимая девушка героя, с которой он вступает в брак.

2. Разумеется, не во всех русских народных волшебных сказках реализуются сразу все формулы. В случае, когда какие-либо из них отсутствуют, происходит перескок по данной строчке последовательности; сказка может начинаться с любой, достаточно далекой от начала формулы. При этом в старых, не «испорченных временем» сказках перескок бывает только вперед: возврата назад по последовательности не происходит.

3. Каждая сюжетная формула может заменяться на противоположную. Скажем, вместо адресованного герою запрета ему может быть дано приказание сделать что-то.

## **2.С. Сюжетная формула как схема и ее конкретизация**

Посмотрим на одном примере, как происходит конкретизация некоторых сюжетных формул. Самая первая формула («отлучка») может осуществляться любым персонажем, по какой-либо причине покидающим родной дом: это принц, отправляющийся на войну; купец, собравшийся в дальний путь; один из родителей, уходящий из дому и наказавший детям — а это и есть «запрет» (формула № 2) — никому не открывать дверь, не дотрагиваться до какого-либо предмета, не входить в потайную комнату и др. Можно указать правила, обеспечивающие при определенных условиях выбор одной или нескольких конкретизаций данной формулы (т. е. выбор персонажа, конкретного действия и т. д.).

Множество таких правил, в свою очередь, может составить алгоритм и программу, позволяющие не только восстанавливать старые сказки, но и строить новые, подобно тому, как, располагая конечным числом нотных знаков музыкальной нотации, мы можем составлять бесконечное количество мелодий.



### III. МАТЕМАТИЧЕСКИЕ УТВЕРЖДЕНИЯ В ХУДОЖЕСТВЕННЫХ ТЕКСТАХ

Читателю, который продрался сквозь тернии материала данной книги и благополучно дошел до этой страницы, мы надеемся доставить удовольствие, обратив его внимание на то, как можно «прозаические» и «скучные» математические положения облечь в изящную поэтическую форму. При взгляде на поэтическое творчество невольно приходит в голову мысль, что поэтов издавна притягивала к себе математика. (Многие из приводимых ниже примеров были собраны Э.П.Казанджаном и приведены в его книге: Школьник — абитуриент — студент — инженер. — М., 1990.)

#### 1. ПОЭТЫ О МАТЕМАТИКЕ

Именно поэтам принадлежат многие образные и вместе с тем исключительно точные высказывания о математике. Вот, например, слова Гете: «Говорят, что цифры правят миром; я знаю одно — цифры показывают, хорошо или плохо он управляется». Еще более высоко оценивает числа Н.Гумилев: «...Потому что все оттенки смысла умное число передает». Многообразные и сильные эмоции по отношению к числам выражает В.Брюсов: «Пред волей числ мы все рабы» и «Вам поклоняюсь, вас желаю, числа!»

Некоторых поэтов, таких, как Велимир Хлебников, числа вообще вели по миру: «Я всматриваюсь в вас, о числа... Вы позволяете понимать века». У него же встречаем такие строки:

Походы мрачные пехот,  
Копьем убийство короля,  
Дождь звезд и синие поля  
Послушны числам, как заход.  
Года войны, ковры чуме,  
Сложил и вычел я в уме,  
И уважение к числу  
Растет, ручьи ведя к руслу.

Сходство и, если угодно, родство душ математика и поэта неоднократно отмечалось. Правда, как это ни удивительно, самые элементарные математические факты и утверждения многим поэтам не даются или даются с трудом. Некоторые поэты, например Е.Винокуров, смело в этом признаются:

Я чуть не плакал. Не было удачи!  
Задача не решалась — хоть убей.  
Условье было трудным у задачи,

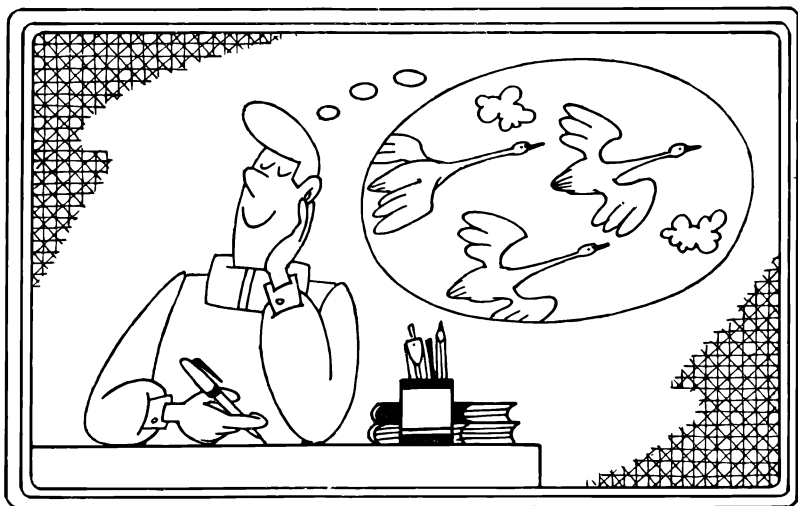
Дано:  
«Летела стая лебедей...»  
Я, щеку грустно подперев рукою,  
Делил, слагал — не шли дела на лад!  
Но, лишь глаза усталые закрою,  
Я видел ясно:  
Вот они летят...

Читая эти строчки, очень хочется пожалеть поэта. Действительно, бывает ли условие труднее, чем «летела стая лебедей...»? Есть отчего щеку грустно подпереть рукою. Конечно, задачу решить отнюдь не просто, если замкнуться в поэтическом мире, и, вместо того чтобы строить математическую модель задачи, закрыть усталые глаза, и представить себе летящих лебедей. Поэтический настрой сказался даже в том, что автор вместо принятого в математике обозначения действия сложения — глагола *складывал* — употребил глагол *слагал*, который более уместен в сочетании *слагать стихи*. Не потому ли и «не шли дела на лад» у поэта?

Полны трагизма и такие строки того же поэта:

Сию сутулясь, предо мной учебник, —  
Моя мольба восходит горячо!..  
...Но разве можно без наличия чуда  
Постичь задачу?!

Образ поэта, который сидит в позе, весьма не полезной для здоровья, перед учебником с мольбой о чуде и в тщетной надежде «постичь задачу», вызывает искреннее сочувствие. В



наше время, когда есть такие большие возможности восполнить пробелы в математическом образовании, посещая различные школы, курсы, занимаясь самообразованием, можно посоветовать поэту избавиться от ужаса, который внушает ему математика, воспользовавшись для этого одной из предоставляемых возможностей.

Читая поэтические строчки, трудно понять, какие разделы математики даются поэтам труднее. У О.Мандельштама познания в области математики, видимо, заканчивались фрагментами таблицы умножения и сложением однозначных чисел. Он же сам писал:

Дважды два — четыре,  
Два да три — пять.  
Вот и все, что мы можем,  
Что мы можем знать.

Впрочем, наверное, эти строчки мы прочли слишком поверхностно. Не скрываются ли в них такие глубокие философские проблемы, как проблема знания и его ограниченности? Вспомним известное изречение Сократа: «Я знаю только то, что ничего не знаю». Ведь знать лишь то, что дважды два — четыре и два да три — пять, — это все равно, что ничего не знать из области математики.

«Дважды два — четыре». Казалось бы, найдется ли более простое математическое утверждение? На ум приходят слова детской песенки:

Дважды два — четыре,  
Дважды два — четыре,  
Это всем известно в целом мире.  
Дважды два — четыре,  
Дважды два — четыре,  
А не три, а не пять. Это надо знать!  
Дважды два — четыре,  
Дважды два — четыре,  
А не шесть, а не семь.  
Это ясно всем!

Есть, однако, поэты, например И.Эренбург, у которых даже с такими утверждениями возникают сложности.

Я не знал, что дважды два — четыре,  
И учитель двойку мне поставил, —

признавался он. Впрочем, что тут удивительного, ведь И.Эренбург сам замечал: «Не был я учеником примерным».

Во Франции, где преподавание математики еще в прошлом веке было на высоком уровне, даже поэтам удавалось овладеть всеми четырьмя действиями над однозначными

числами. Но двухзначные числа для некоторых поэтов все же явились камнем преткновения. Путается в них даже такой замечательный поэт, как Жак Превьер:

Четыре и четыре — восемь,  
восемь и восемь — шестнадцать,  
шестнадцать и шестнадцать — сколько,  
сколько в сумме они составляют?

Повторение слова *сколько* говорит о мучительных поисках ответа, который, судя по последней строчке, так и не был найден.

## 2. МАТЕМАТИЧЕСКИЕ ОБРАЗЫ В ПОЭЗИИ

Посмотрим теперь, как у поэтов обстоят дела с геометрией и тригонометрией. Наверное, должен знать геометрию окончивший архитектурный институт А. Вознесенский. Не случайно он назвал один из своих поэтических сборников «Треугольная груша». В скобках заметим, что А. Вознесенский знаком со свойствами и графиками основных элементарных функций, в частности с параболой: «Судьба, как ракета, летит по параболе».

У других поэтов ярко выражено свое субъективное отношение к разным геометрическим объектам. Широко известны строки П. Когана:

Я с детства не любил овал.  
Я с детства угол рисовал.



А у Н.Коржавина вкус совсем иной:

Я с детства полюбил овал  
За то, что он такой законченный.

Не следует, однако, думать, что геометрия поэтам дается легко. Вот какие неприятности причиняли геометрия с тригонометрией А.Поперечному:

Я постигал с трудом  
Неясный смысл  
Великих чисел и простых затей.  
Но часто  
Над руками коромысел  
Мне снился треугольный журавель.

И, сложные фигуры начертая,  
Зубря тригонометрию  
В зарю,  
Я говорил себе:  
А на черта я  
Ее, тригонометрию, зубрю?!

Этому же поэту плохо дается действие извлечения корня:

И презирая точные науки  
(В них коренное коренилось зло),  
Я корень извлекал,  
Идя на муки,  
Не извлекая истины зерно.

Не скрывает своих эмоций по поводу разнообразных разделов математики поэт И.Снегова:

Математика — это трудно.  
Это дар. С первых лет. От бога.  
Слишком промахи в ней подсудны.  
Слишком взыскивает с итога.  
Уравнения, в которых скопом  
Корни, степень, неравенств бездна.  
Суть, замкнувшаяся по скобкам,  
И — до дьявола неизвестных.  
Или дробь... Ох, эти дроби!

И после всех этих признаний не удивительно, что автор заключает:

Ни одно из моих решений  
Не сходилось вовек с ответом.

Для Е.Евтушенко едва ли не вершиной математической сложности является заурядное школьное понятие логарифма (правда, изучаемое лишь в старших классах). В одном из стихотворений он использует это понятие как эквивалент сложности: «...Но это посложнее логарифма».

Число подобных примеров можно сколь угодно увеличивать. Однако мы хотели бы сейчас показать, что плохое отношение к математике — это отнюдь не черта, свойственная всем поэтам и поэзии в целом.

### **3. ПОЭТИЧЕСКОЕ ОБЫГРЫВАНИЕ МАТЕМАТИЧЕСКИХ ПОНЯТИЙ**

Внимательно вчитываясь в некоторые поэтические строки, можно обнаружить, что в них как бы заново открываются математические закономерности, уточняются разнообразные математические факты и проводятся тончайшие математические рассуждения. Если еще учесть, что все это облечено в изысканную форму, то создается впечатление, что математику можно учить прямо по «поэтическим учебникам».

#### **3. А. Пустое множество**

В разделе, где шла речь о таком важном понятии, как множество, мы ввели понятие пустого множества, объяснив, что при любом описании пустое множество оказывается одним и тем же — число элементов в нем равно нулю. Это обстоятельство С.Маршак иллюстрирует следующим образом:

Спросил меня голос в пустыне дикой:  
— Много ли в море растет земляники?  
— Столько же, сколько селедок соленых  
Растет на берегах и елках зеленых.

#### **3. В. Прямая и обратная пропорциональность**

Известное понятие прямой и обратной пропорциональности очень близко поэтам.

Вот, например, как пишет П.Вяземский:

Чтоб более меня читали,  
Я стану менее писать.

Вспомним также известные строки А.Пушкина:

Чем меньше женщину мы любим,  
Тем легче нравимся мы ей  
И тем ее вернее губим  
Средь обольстительных сетей.

#### **3.С. Предел**

Даже такие сложные математические понятия, как предел функции, находят отражение в поэзии. Более того, некоторым поэтам бывает хорошо известен тот факт, что не любая функция имеет предел. Приведем в качестве примера четыре строки М.Лермонтова:

Как я хотел себя уверить,  
Что не люблю ее, хотел  
Неизмеримое измерить,  
Любви безбрежной дать предел.

И две строчки В.Брюсова:

Люблю в мечтах предел.  
Меня страшит безмерность...

На первый взгляд понятие предела не должно было быть известно Е.Евтушенко, ведь, разумеется, это понятие гораздо сложнее, чем логарифм. Однако поэтические пути неисповедимы: предел и связанные с ним математические факты оказываются поэту знакомыми. Даже, считая поначалу, что предел есть у всякой функции, он тут же спохватывается, самостоятельно обнаружив опровергающий пример (попутно заметим, что утверждения со словом *все* могут быть, как мы видим, опровергнуты одним-единственным противоречащим примером):

Предел на белом свете есть всему:  
любви, терпенью, сердцу и уму,  
и мнимой беспредельности простора.  
... Есть у любого гения предел -  
лишь подлость человечья беспредельна.

### 3.D. Кванторы

О так называемых кванторах и кванторных словах мы подробно расскажем в восьмой главе, а пока попробуем посмотреть, как поэты употребляют кванторные слова. Кстати, подчеркнем, что сознательное употребление кванторов в математике в их современной интерпретации началось лишь сравнительно недавно. Между тем поэты, видимо, издавна глубоко ощущали разницу между квантором общности и квантором существования. Более того, для выражения нужных им кванторных смыслов они ставили кванторы в требуемом законами языка порядке. Вот пример из В.Блейка:

Каждый день на белом свете  
Где-нибудь родятся дети.

Здесь мы видим два кванторных слова *каждый* и *где-нибудь*. Первое слово — это типичный квантор общности, а второе скрывает в себе квантор существования: *где-нибудь* означает 'в каком-нибудь месте'. Логическая структура этой поэтической формулы такова: 'Для всякого дня существует (на белом свете) место, такое, что...'

Прочитаем короткое двустишие Н.Заболоцкого:

Как хорошо, что дырочку для клизмы  
Имеют все живые организмы.

Порядок следования языковых кванторных слов (смыслов) не совпадает здесь с порядком следования кванторов в формальной логической структуре данного утверждения. Эта

структура такова: '... У всякого живого организма и м е е т с я (существует) дырочка для клизмы'.

Следующая эпиграмма С.Леца по логической структуре сходна с утверждением В.Блейка:

Для любого осла  
Где-то травка проросла.

#### 4. ЛОГИЧЕСКИЕ РАССУЖДЕНИЯ В ХУДОЖЕСТВЕННЫХ ТЕКСТАХ

##### 4.А. Доказательство от противного

Логические рассуждения многих поэтов четки и отшлифованы. В следующей эпиграмме П.Сумарокова блестяще проведено доказательство от противного:

Что Клав меня лечил, слух этот, друг мой, лжив:  
Когда б то было так, то как же б я был жив?

Разберем это поэтическое доказательство, выявляя его логическую структуру. Как известно, доказательства от противного строятся по некоторой общей схеме, имеющей две модификации:

1) Требуется доказать, что утверждение *A* истинно.

Предположим, что *A* ложно. Тогда получаем, что некоторое *B* ложно.

Между тем про *B* заранее известно, что оно истинно.

Следовательно, наше предположение было неверным, и потому *A* истинно.

2) Требуется доказать, что утверждение *A* ложно.

Предположим, что *A* истинно. Тогда получаем, что некоторое *B* ложно.

Между тем про *B* заранее известно, что оно истинно.

Следовательно, наше предположение было неверным, и потому *A* ложно.

В приведенной эпиграмме реализуется вторая модификация общей схемы. Приведем полностью логическую структуру эпиграммы: 'Требуется доказать: утверждение («слух»), что Клав лечил автора, ложно.

Предположим, что этот «слух» истинен. Тогда получили бы, что автор умер (то есть ложно, что он жив).

Между тем, по-видимому, нет сомнения, что разбираемые строчки написаны автором в тот момент, когда он был жив.

Следовательно, наше предположение, что Клав лечил автора, неверно'.

Аналогичную схему рассуждений можно усмотреть в отрывке из пушкинского «Моцарта и Сальери»:

М о ц а р т : Ах, правда ли, Сальери,  
Что Бомарше кого-то отравил?



И сам же Моцарт опровергает это предположение, основываясь на следующем рассуждении:

Он же гений  
Как ты да я. А гений и злодейство  
Две вещи несовместные.

Попробуйте сами восстановить все необходимые звенья этого доказательства от противного.

Пытаясь доказать, что высокая музыка не нуждается в словах, Н.Матвеева также использует схему «от противного»: 'Если бы музыка объяснялась словами, то она состояла бы из слов, а не из музыкальных звуков':

Вы объяснили музыку словами.  
Но, верно, ей не надобны слова —  
Не то она, соперничая с вами,  
Словами изъяснилась бы сама.  
И никогда (для точности в науке),  
Не заводя пустых обиняков,  
Не тратила бы времени на звуки,  
А сразу начинала бы со слов.

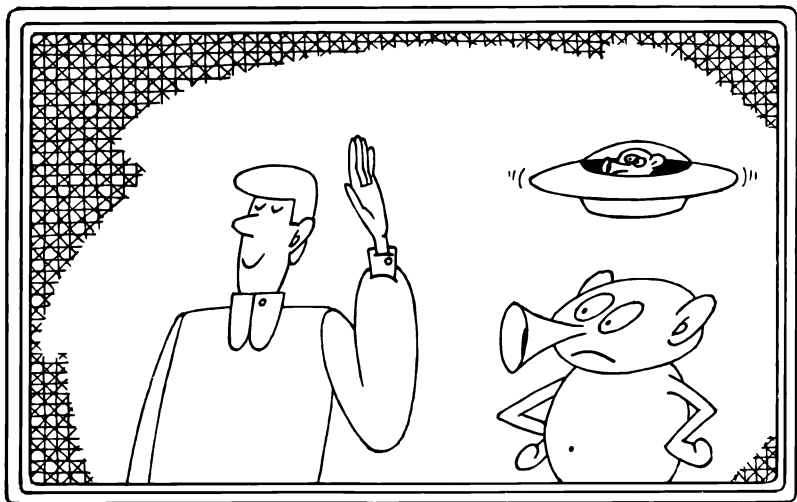
#### 4.В. Модальные рассуждения

Поэты зачастую используют и более сложные логические схемы рассуждений, изучаемые математиками в таком трудном разделе математической логики, как «Модальная логика». Важным компонентом всех модальных рассуждений являются модальные слова, выражающие значение возможности и необходимости: *может, возможно, необходимо* и др.

Рассуждения, укладываемые в схемы модальной логики, разнообразны по своей структуре. Более полный перечень типов модальных рассуждений с иллюстрациями из художественной литературы можно найти в статье: Булыгина Т.В., Шмелев А.Д. Возможности естественного языка и модальная логика//Язык логики и логика языка. — М., 1990. Ограничимся лишь несколькими примерами.

Одним из наиболее распространенных видов модальных рассуждений является рассуждение «от действительного к возможному», имеющее следующую структуру: «Если *А* есть на самом деле, то *А* возможно». Вот отрывок, принадлежащий перу замечательного русского поэта М.Цветаевой: «Пушкин из всех женщин на свете любил свою няню, которая была не женщина. Из «К няне» Пушкина я на всю жизнь узнала, что старую женщину можно любить больше, чем молодую».

Сначала несколько слов о самом тексте Цветаевой. Кажущаяся парадоксальность вызвана тем, что поэт использует сочетание *не женщина* в значении 'старая женщина'. На это указывает заключительная фраза приводимого отрывка.



Уяснив смысл, мы можем вскрыть логику умозаключения поэта: 'Пушкин любил няню (больше, чем всех остальных женщин). Няня Пушкина — старая женщина. Следовательно, старую женщину можно любить (больше, чем молодую)'. Утверждения *Пушкин любил няню* и *Няня — старая женщина* относятся, как говорят логики, к модусу действительности, а заключение — к модусу возможности.

В бытовом языке невозможным часто называется то, что лишь казалось невозможным. Нередко встречающаяся фраза *Случилось невозможное* не понимается буквально, а означает, что произошло нечто неожиданное. Но по законам модальной логики, коль скоро нечто произошло (иными словами, имеет место в действительности), то оно, в силу схемы «от действительности — к возможности», является возможным. Этим объясняется парадокс во фразе А.Блока *И невозможное возможно*.

В ответ на некоторое сообщение можно услышать реплики типа *Не может быть! Быть того не может!* Разумеется, такие реплики направлены не на то, чтобы обидеть собеседника недоверием, а выражают удивление и заинтересованность услышанным. Но чуткая к логике М.Цветаева, тонко ощущая игру смыслов, в ответ на первую реплику восклицает: «Может, раз есть!», явно обращаясь к схеме «от действительности — к возможности».

В ситуации, когда мы непосредственно наблюдаем, что *А* имеет место, наши отвлеченные представления о якобы невозможности *А* теряют силу в соответствии все с той же

схемой «от действительности — к возможности». Правда, на практике люди склонны отвергать очевидное, основываясь лишь на своих откуда-то взявшихся представлениях о невозможности той или иной ситуации. Пародируя нечуткость и невосприимчивость таких людей к реально наблюдаемым фактам, можно сказать как бы вслед за ними: «Этого нет, потому что не может быть никогда».

Иное дело, когда невозможность изначально задана языком. Например, она может вытекать прямо из значений слов, употребляемых носителями языка. Такую невозможность нельзя опровергнуть ни умозаключениями, ни даже наблюдаемыми фактами. Как всегда, поэты облачают эту мысль в художественную форму:

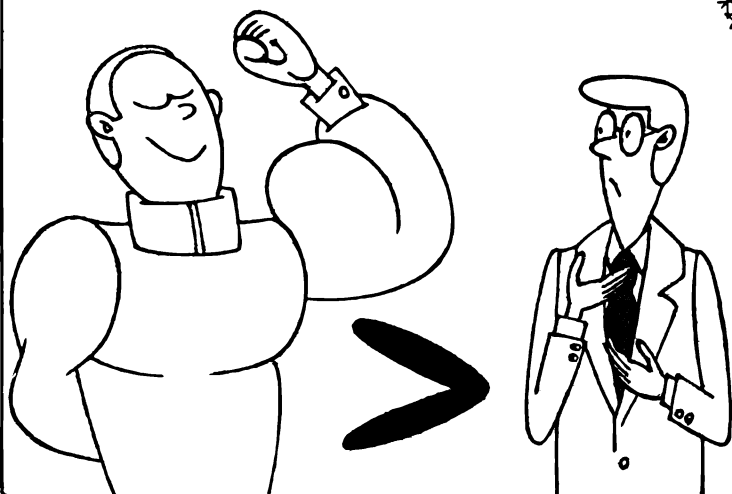
Мятеж не может кончиться удачей,  
В противном случае зовут его иначе.

*(Д.Харрингтон, пер. С.Маршака)*

Читатель, конечно, понял, что за нашим полушутливым, полусерьезным тоном скрыт анализ многих очень важных как для математики, так и для языкознания формальных положений и фактов. Наша цель была не только в том, чтобы обсудить сами эти факты и связанные с ними следствия, но и показать принципиальную возможность облечь серьезные размышления в несерьезную форму. Ведь раз мы это сделали — значит, это возможно!

Глава восьмая

**Языковой анализ  
математических  
текстов**



## 1. ЯЗЫКОВОЙ ПОДХОД К РЕШЕНИЮ ЗАДАЧ

**Ч**то нужно для того, чтобы правильно решить математическую задачу? «Станный вопрос, — скажете вы, — это же ясно: знать математические понятия и связи между ними, т.е. всякие там аксиомы, теоремы; ну, может быть, еще знать приемы решения задач». Мы хотим сразу же сказать, что всего этого недостаточно. Решение задач требует не только собственно математических навыков и умений, но и определенной языковой культуры, хорошего владения естественным языком. В частности, речь идет об умении подойти к тексту задачи (или какому-то другому математическому тексту) как к особому типу текста на естественном языке, т.е. фактически решать традиционную лингвистическую проблему, называемую анализом текста. Иными словами, человек, приступающий к решению задачи, должен не только видеть в ней математическую проблему, но и относиться к ней как к языковому материалу, требующему лингвистической расшифровки.

Более того, в решении задач (например, арифметических или на составление уравнений) лингвистическая деятельность, как правило, предшествует математической. Действительно, непосредственно к математической деятельности относится сам процесс решения уравнений, неравенств и т.д., тогда как способ их получения составляет не только собственно математическую проблему, но и в значительной мере связан с лингвистическим разбором текста поставленной задачи. Таким образом, изучение языковых фактов оказывается как бы встроенным в математическую науку. Можно было бы даже сказать, что при прочих равных условиях чем лучше человек знает языковые законы и умеет работать с текстом на естественном языке, тем лучше он умеет решать математические задачи.

Недооценка роли языковых фактов мешает понять условие задачи (например, приводит к тому, что она оценивается как некорректно составленная, непонятная, странная или логически противоречивая). Вот лишь один простейший пример: «Сумма двух чисел равна 21. Две пятых одного числа и две трети другого равны 10. Найти эти числа». У неподготовленного в языковом отношении ученика вполне может вызвать недоумение употребление союза *и*. Как здесь его понимать? Если две пятых одного числа равны 10 и две трети другого тоже равны 10, то легко подсчитать, что первое число равно 25, а второе 15. Но ведь в условии говорится, что сумма

этих чисел равна 21. Выходит, что при таком понимании задача поставлена некорректно, ее условие содержит противоречие. Логичнее предположить, что задачу следует прочесть и понять как-то по-другому. В математических текстах союз *и* часто выступает как языковой шифр операции сложения. В этом случае если *и* является языковым аналогом сложения, то все встает на свои места. Получаем следующий, «правильный» смысл задачи: «Сумма двух пятых одного числа и двух третей другого равна 10». При таком понимании задача уже не содержит никаких противоречий.

Ниже мы покажем некоторые приемы лингвистического анализа математического текста (в основном — текста задач) и те трудности, которые возникают при незнании приемов или неумении ими пользоваться. При этом мы хотели бы дать вам несколько практических советов, облегчающих подход к таким текстам. Естественно, что для этих целей удобнее всего рассмотреть так называемые «текстовые» задачи.

## II. АРИФМЕТИЧЕСКИЕ ЭЛЕМЕНТЫ В МАТЕМАТИЧЕСКОМ ТЕКСТЕ

### 1. ПРИМЕРЫ АРИФМЕТИЧЕСКИХ ЭЛЕМЕНТОВ

Прежде всего в тексте следует выделить те языковые фрагменты, которые являются ключевыми при составлении уравнений (систем, неравенств и др.), т. е. при построении математической модели задачи (о понятии модели речь шла в главе 3). Такие фрагменты мы будем называть *а р и ф м е т и - ч е с к и м и* элементами или арифметическими словами, сочетаниями слов и т. д.

#### 1.А. Имена чисел

Приведем примеры некоторых арифметических элементов.

Во-первых, это имена чисел. Их легко обнаружить, когда они выражены количественными числительными в цифровой или словесной записи. Однако обозначения чисел бывают и более сложными. Мы можем встретить в тексте такие описательные выражения, как *первоначальная цена костюма; остальные книги; общее число автомобилей, изготовленных за месяц*. Эти выражения в модели задачи заменяются буквами, а соответствующие количественные значения либо сообщаются в условии задачи, либо, если это необходимо, определяются в процессе ее решения.

Важно, чтобы разным в смысловом отношении описательным выражениям соответствовали разные буквы, а синонимичным — одна и та же буква. А потому, хотим мы этого

или нет, нам приходится решать отнюдь не тривиальную (и в общем виде еще не решенную) лингвистическую проблему установления смысловой эквивалентности двух языковых единиц в составе текста. Когда выражения достаточно просты по своей структуре, эту проблему синонимии можно решать сравнительно простыми средствами, которые могут быть формально заданы и реализованы на ЭВМ. Однако если обратиться к таким выражениям, как *количество самолетов, которые необходимо выпустить по плану = плановое задание — план*, то для распознавания их смысловой эквивалентности требуются уже достаточно сложные мыслительные операции, которые не так-то просто запрограммировать. Может быть, здесь и лежит одна из причин неудач в создании алгоритмов автоматического решения задач.

## 1.В. Языковые выражения и символика

В качестве еще одного примера важных арифметических элементов можно привести языковые выражения и символы, обозначающие арифметические операции и отношения. В модели им соответствуют определенные обозначения: знаки  $+$ ,  $-$ ,  $:$ ,  $=$ ,  $<$ ,  $>$  и т. п. Обратим внимание читателей на некоторые «подводные камни», которые могут привести к ошибкам при построении модели. Все эти ошибки носят ярко выраженный языковой характер.

## 2. МАТЕМАТИЧЕСКАЯ И ЯЗЫКОВАЯ МОДЕЛЬ РЕШЕНИЯ ЗАДАЧ

### 2.А. Арифметические и неарифметические слова

Конструирование математической модели бывает затруднено из-за того, что одно и то же слово в одной задаче может быть арифметическим, а в другой — неарифметическим. Сравним тексты двух задач: а) Скосили  $\frac{3}{7}$  всего луга. Найдите площадь всего луга, если скосили 23 га. б) Туристы проехали поездом 840 км и теплоходом 415 км, после чего им осталось проехать на автомашинах 265 км и пройти пешком 72 км. Определите длину всего маршрута.

Нас будет интересовать слово *весь* (*всего*), которое входит в оба текста. Отметим, что функции данного слова в этих текстах совершенно разные. В первом тексте оно не служит указанием на то, как надо решать задачу. В таких случаях говорят о чисто стилистических функциях слова (говорят, что с этим словом текст лучше звучит, легче читается и т. п.). Во второй задаче слово *весь* является арифметическим. Его назначение — указать, что к ответу следует прийти, выполнив сложение. Иными словами, *весь* заменяет слово *сумма* — ведь речь идет о сумме длин маршрутов, указанных в условии

задачи. Опустить арифметическое слово в тексте, вообще говоря, нельзя, без него текст становится непонятным.

## 2. В. Разные понимания арифметических слов

Более сложной является ситуация, когда в каждом из двух контекстов слово является арифметическим, но имеет в этих контекстах разные значения. Возьмем для примера слово *больше*. В одних случаях оно выступает как показатель отношения «больше», а в других — называет операцию (сложение или вычитание) одновременно с ее результатом. Фраза *Число  $a$  больше числа  $b$*  переводится на язык математических символов как  $a > b$ , а фраза *Число  $a$  больше числа  $b$  на  $c$*  — как  $a - b = c$  или  $a = b + c$ . Впрочем, строго говоря, фразе *Число  $a$  больше числа  $b$  на  $c$*  соответствует не равенство  $a - b = c$ , а система

$$\begin{cases} a - b = c, \\ c > 0, \end{cases}$$

так что смысл отношения 'больше' присутствует не только в первом случае, но и во втором. Различие двух приведенных фраз, однако, существенно: в первой смысл 'больше' — основной и единственный, а во второй — дополнительный (возможно даже, что в первой фразе сказуемое — слово *больше*, а во второй — *больше на*). Более того, если  $a$  и  $b$  получают конкретные числовые значения, то запись  $c > 0$  избыточна и потому опускается: фразе «*Число 20 больше 15 на 5*» в математической модели достаточно поставить в соответствие равенство  $20 - 15 = 5$ ; тот факт, что  $5 > 0$ , тривиален и в модели обычно не фиксируется.

Советуем обратить внимание на то, что слово *больше* в сочетании *больше или равно* понимается всегда только в первом смысле, т. е. указывает, как и слово *равно*, на отношение между числами. Именно поэтому нельзя сказать *больше или равно на...*: слово *на* заставляет интерпретировать *больше* как операцию.

Такой же смысловой двойственностью обладают слова *меньше*, *превосходит* и ряд других, в то время как выражения *не превосходит*, *меньше или равно* однозначно понимаются только как названия отношений.

Уже эти сравнительно простые примеры показывают роль языковой информации как при анализе текстов задач, так и при их составлении.

При написании текста задачи ее автор выступает по меньшей мере в двух функциях: редактора-стилиста и редактора-математика. Если автор задачи неудачно употребил неарифметическое слово, то это еще не говорит о том, что он плохой математик, хотя явно свидетельствует о его не вполне удовлетворительном владении языком. В то же время



неверное употребление арифметических слов демонстрирует наличие у него каких-то пробелов как в языковой, так и в математической подготовке.

### III. ЛОГИЧЕСКИЕ ЭЛЕМЕНТЫ В МАТЕМАТИЧЕСКОМ ТЕКСТЕ

#### 1. КВАНТОРЫ И КВАНТОРНЫЕ ВЫРАЖЕНИЯ

При подходе к задаче очень важно устанавливать наличие в ее тексте слов или особых смыслов, которые в лингвистике принято называть соответственно кванторными словами и кванторными смыслами. Остановимся на них подробнее.

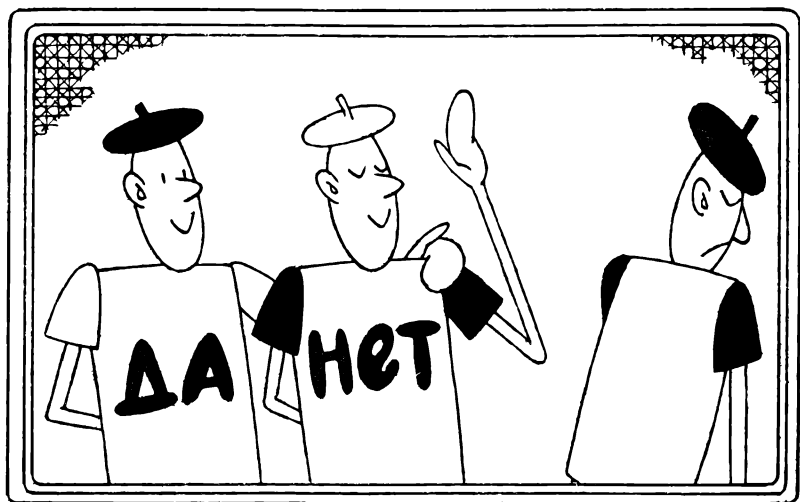
##### 1. А. Квантор общности и квантор существования

В языке математики основными кванторными единицами являются квантор общности и квантор существования. Последние являются также элементами тех формальных языков, которые строятся лингвистами или математиками для более точного описания смысла предложений естественных языков. Не вдаваясь в сложную проблему анализа кванторных слов и смыслов естественного языка и их соответствия кванторам языка математики, ограничимся здесь отдельными замечаниями и обсуждением ряда примеров. Читателю, желающему поподробнее ознакомиться с тем, что такое кванторы, узнать, каковы формальные способы обращения с ними, мы рекомендуем заглянуть в учебники или популярные книги по математической логике (*Гусев В.А., Орлов А.И., Розенталь А.Л.* Внеклассная работа по математике в 6—8 классах. — М., 1984; *Балк М.Б., Балк Г.Д.* Математика после уроков. — М., 1971. — Глава 7, § 2; *Кальбертсон Дж.* Математика и логика цифровых устройств. — М., 1965; *Шиханович Ю.А.* Введение в современную математику. — М., 1965).

В первой из этих книг (на с. 22—23) рассматривается следующий пример. Представим себе, что в ящике могут лежать белые и красные шары. Все шары одинакового размера и неразличимы на ощупь. Кто-то вынимает шар из ящика. Шар оказывается, к примеру, красным, и мы кладем его обратно. Можно ли сделать какое-то заключение о цвете шаров в ящике?

Неправильно было бы сказать: «Все шары красные». Верно лишь то, что красными являются н е к о т о р ы е (но не обязательно все) шары. Может быть даже так, что вынутый шар вообще единственный красный среди шаров в ящике.

В такой ситуации заведомо неверно утверждение, что все шары белые. Верны же его отрицание «Не все шары белые»



или утверждение «Некоторые (по крайней мере один) шары красные».

**З а м е ч а н и е.** Здесь мы сталкиваемся с важным логическим законом: верно либо утверждение, либо его отрицание. В логике этот закон называется законом исключенного третьего.

Так, если утверждение «Все предметы из  $M$  обладают свойством  $A$ » (или, что то же самое, «Любой предмет из  $M$  обладает свойством  $A$ ») верно, то неверно его отрицание: «Хотя бы один предмет из  $M$  не обладает свойством  $A$ » (или, что то же: «Существует предмет из  $M$ , не обладающий свойством  $A$ »). И наоборот, если верно отрицание, то неверно само утверждение.

Утверждения обо всех объектах в формальном языке записи предложений содержит выражение, называемое **к в а н т о р о б щ н о с т и**, а утверждения лишь о некоторых объектах — выражение, называемое **к в а н т о р с у щ е с т в о в а н и я**. Таким образом, квантор общности является математическим аналогом слов типа *все, любой* и т. д., а квантор существования — слов *некоторые, существует, хотя бы один* и подобных.

Квантор общности, как мы видели, связан с квантором существования через операцию отрицания. Отрицание предложения с квантором общности содержит квантор существования и наоборот.

Убедитесь в этом, построив отрицание следующих утверждений: 1) Все люди смертны; 2) Для любого  $x$  из

множества  $R$   $x^2 \geq 0$ ; 3) Все простые числа, больше двух, нечетны; 4) Хотя бы для одного  $x$  из множества  $R$   $x^2 - 2x + 1 = 0$ .

## 1. В. Кванторные слова и кванторные смыслы в русском языке

В русском языке кванторные смыслы обычно передаются словами *все, каждый, всякий, любой, какой бы то ни было* и другими (для квантора общности); *существует, некоторые, есть, имеется, надеется, отыщется* (для квантора существования). Разумеется, внутри каждой из двух групп слов между отдельными словами могут быть тонкие языковые и, в частности, смысловые различия (например, можно сказать *Все люди братья*, но нельзя сказать *Каждый человек брат*), но этими различиями как раз в большинстве случаев и можно при решении математических задач пренебрегать. А вот различием между квантором общности и квантором существования, как мы уже показали, пренебрегать нельзя.

Разбор текста особенно сложен, когда в нем кванторы соседствуют. Например, порядок слов, при котором квантор общности предшествует квантору существования, обычно выражает иной смысл, чем порядок, при котором квантор существования предшествует квантору общности. Так, предложение *В каждой очереди есть человек, пытающийся пролезть без очереди* не синонимично предложению *Есть человек, пытающийся пролезть без очереди в каждой очереди*, а предложение *Любой человек является членом некоторого коллектива* предложению *Некоторый коллектив имеет своим членом любого человека*. Больше того, может быть так, что при первом порядке математическое утверждение истинно, а при втором — ложно. Например, утверждение *Для любых трех точек, не лежащих на одной прямой, существует точка, равноудаленная от них* выражает истинную теорему из курса геометрии (теореме о существовании центра окружности, описанной около треугольника, равносильную известной теореме о существовании окружности, описанной около треугольника). Попробуем теперь переставить кванторные выражения: *Существует точка, равноудаленная от любых трех точек, не лежащих на одной прямой*. Очевидно, что это утверждение ложно.

Построение отрицания для предложений с несколькими кванторами часто представляет собою весьма нетривиальную задачу. Она предполагает умение обращаться не только с кванторами в формальном языке (т.е. свободное владение математическими понятиями и математическим языком), но и с кванторными словами естественного языка (т.е. хорошее владение естественным языком). Вот попробуйте, например, решить такую задачу:

*Назовем контрольную легкой, если в каждом классе найдется такая парта, что всякий ученик, сидящий за ней, может справиться с этой контрольной. Дайте определение трудной контрольной (предполагается, что контрольные бывают либо трудными, либо легкими).*

В нормальной ситуации порядок следования кванторных слов в тексте соответствует порядку следования кванторов в формальном логическом языке. Но это бывает не всегда. Например, в тексте задачи *Всякое сечение некоторого тела есть круг. Доказать, что это тело — шар* порядок слов *всякий — некоторый* означает 'Существует тело, такое, что всякое его сечение есть круг' (порядок *существует — всякий*).

Подытоживая сказанное о кванторах и кванторных словах, можем дать следующий совет: при анализе текстов аксиом, теорем и других утверждений следует внимательно отнестись к кванторным выражениям и порядку их следования во фразе. Как показывает опыт авторов, многие ошибки при построении истинных утверждений и решений уже готовых задач объясняются не плохой математической подготовкой, а невнимательным отношением к кванторным словам естественного языка.

### **1.С. Явные и неявные кванторные смыслы**

Трудности усугубляются, когда кванторный смысл, несомненно присутствующий во фразе, не может быть приписан ни одному слову в ней. В таких случаях лингвисты говорят о неявных кванторных смыслах.

Рассмотрим два известных вам предложения:

*В прямоугольном треугольнике сумма квадратов катетов равна квадрату гипотенузы.*

*Две хорды, которые выходят из одной точки окружности и концы которых совпадают с концами диаметра той же окружности, являются катетами прямоугольного треугольника.*

Ясно, что в первой фразе речь идет о всяком прямоугольном треугольнике, а во втором — о каком-то, о некотором. Между тем в явном виде кванторные слова и в первом, и во втором предложении отсутствуют.

Неумение распознавать «скрытый» в предложении квантор может привести к неверному решению задачи или неточному ее пониманию, а неуместное опущение кванторных слов при построении математического текста нередко приводит к некорректным формулировкам. Рассмотрим следующее предложение: *В прямоугольном треугольнике два катета равны между собой.* Как его понимать? Как оценить его с точки зрения истинности-ложности?

Оно будет истинно, если выражение *в прямоугольном*

*треугольнике* понимается как содержащее неявный квантор существования ('в некотором прямоугольном треугольнике'). Возьмем, например, текст следующей задачи: *В прямоугольном треугольнике два катета равны между собой. Найти углы этого треугольника.* Ясно, что условие этой задачи не содержит никаких противоречий, ведь прямоугольные треугольники могут быть равнобедренными. Если бы рассматриваемое предложение было ложным, оно не могло бы быть условием задачи.

Но то же самое предложение уже будет ложным, если понять его как *Во всяком прямоугольном треугольнике...*, т.е. как содержащее скрытый квантор общности. Сформулированное как теорема, данное утверждение, как и утверждения, близкие ему по структуре (вспомним приведенную чуть выше формулировку теоремы Пифагора), скорее всего, будет понято именно таким образом. Вот почему учитель, если он услышит от вас такую формулировку «теоремы», сочтет ее ложной со всеми вытекающими отсюда последствиями. Учитывая важность и языковую сложность кванторных слов, советуем вам не опускать их совсем или опускать с большой осторожностью.

### 1.D. Кванторные и некванторные значения слов

Одно и то же слово в разных предложениях может иметь разные значения: кванторное и некванторное. Возьмем, например, слово *все*. Мы говорили, что оно может иметь значение квантора общности. Действительно, в утверждении *Все натуральные числа, делящиеся на 4, делятся на 2* или в вопросе *Все ли простые числа нечетны?* слово *все* выступает в функции квантора общности. Между тем в предложении *Областью определения данной функции являются все натуральные числа* слово *все* имеет более сложный смысл, а в вопросе *Сколько деталей изготовили все рабочие цеха?* слово *все* вообще не выражает кванторного значения: оно выступает как арифметическое слово-операция. В математической модели слову *все* в данном случае, скорее всего, соответствует сложение: для решения задачи требуется сложить количество деталей, изготовленных каждым рабочим в отдельности. Из последнего примера ясно, что определение функций слова *все* в исходной формулировке условия задачи может служить ключом к ее решению.

### 2. НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ

Большие трудности в понимании условий разнообразных задач возникают в связи со словами *может, должен, необходимо, достаточно*, союзами *и, или, если... то, тогда и только*

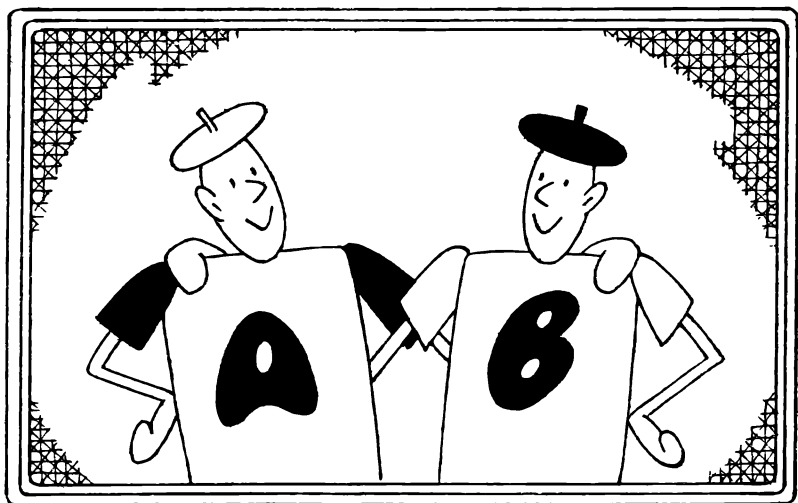
тогда. Выше мы уже видели возможность неоднозначного понимания союза *и* и последствия, к которым может привести неверное понимание этого слова в текстах задач.

Здесь мы остановимся еще на одном вопросе, традиционно трудном для изучающих математику. Речь идет о *необходимых и достаточных* условиях и некоторых языковых особенностях их формулировок. Пренебрежение этими особенностями может повлечь за собой неправильную интерпретацию условия задачи или неверное осмысление теоремы.

## 2.А. Необходимо и достаточно

В математических текстах утверждение «*A* — достаточное условие для *B*» означает то же самое, что и утверждение «*B* — необходимое условие для *A*», а именно то, что во всех случаях, когда справедливо *A*, справедливо и *B*. Например, делимость числа на 4 — достаточное условие для деления числа на 2. Иными словами, во всех случаях, когда справедливо утверждение, что число делится на 4, будет справедливо и то, что число делится на 2.

В тех случаях, когда *A* — достаточное условие для *B*, а *B* — достаточное условие для *A* (или, что то же самое, *B* — необходимое условие для *A*, а *A* — необходимое условие для *B*), употребляют сочетание слов *необходимо и достаточно*: утверждение «*A* необходимо и достаточно для *B*» означает, что из справедливости *A* следует справедливость *B* и наоборот.



Отсюда ясно, что утверждение «*A* необходимо и достаточно для *B*» означает то же, что утверждение «*B* необходимо и достаточно для *A*». Например, *Для того чтобы треугольник был прямоугольным, необходимо и достаточно, чтобы квадрат одной его стороны был равен сумме квадратов двух других сторон; а для того чтобы квадрат одной стороны треугольника был равен сумме квадратов двух других сторон, необходимо и достаточно, чтобы треугольник был прямоугольным.*

Когда в математическом тексте явно употребляются слова *необходимо*, *достаточно* или *необходимо и достаточно*, трудностей с выявлением условий, как правило, не возникает. Однако наряду с этими словами язык знает и другие средства, с помощью которых формулируются необходимые, достаточные или необходимые и достаточные условия.

## 2.В. Союз *если*

Одним из таких языковых средств является союз *если*. В математических текстах часто встречаются предложения вида *Если A, то B*; *B, если A*. В этих случаях возможна путаница, связанная с двояким пониманием фраз такого типа. В утверждении *Если  $x > 1$ ,  $x^2 > 1$*  условие  $x > 1$  является достаточным условием для  $x^2 > 1$ , а в примере *Целое число делится на два, если квадрат его делится на два* условие «целое число делится на два» является необходимым и достаточным для того, чтобы квадрат его делился на два. Таким образом, в текстах теорем и определений вида *Если A, то B* или *B, если A* условие *A* иногда понимается только как достаточное, а иногда — как необходимое и достаточное (см. об этом также в книге: *Шиханович Ю.А. Введение в современную математику. — С. 36—37*). Чтобы избежать такой языковой неоднозначности, при формулировке необходимых и достаточных условий прибегают к оборотам *в том и только в том случае*; *тогда и только тогда*; *если и только если*. Однако эти обороты используют нерегулярно, вследствие чего к предложениям с союзом *если* следует подходить с особой осторожностью при выявлении необходимых и достаточных условий.

Практика показывает, что особую сложность для учащихся представляют вопросы типа: «Сформулируйте достаточное (необходимое) условие того, что некоторое утверждение *A* верно». Например: «Сформулируйте достаточное условие делимости числа на 4» или «Сформулируйте необходимое условие делимости числа на 24». Теперь мы можем сформулировать общую форму ответов на подобные вопросы. Если нужно сформулировать достаточное условие для *A* и используется союз *если...*, то само *A* надо поставить в главное предложение после слова *то* и затем искать условие *B*, такое, что

предложение *Если  $B$ , то  $A$*  будет истинным. Наоборот, если требуется найти необходимое условие для  $A$ , используя тот же союз, то ставим  $A$  в придаточную часть после союза *если* и ищем какое-нибудь математически содержательное (нетривиальное) следствие  $B$ , чтобы утверждение *Если  $A$ , то  $B$*  было истинным.

По-другому ведет себя слово *должен*. Оно может указывать как на необходимое условие, так и на необходимое и достаточное. В рассуждении *Число  $x$  делится на 4, значит, оно должно делиться на 2* слово *должен* говорит о том, что делимость на 2 есть необходимое условие делимости числа на 4. А вопрос *Какому условию должен удовлетворять параллелограмм, чтобы быть квадратом?* предполагает ответ в виде указания необходимого и достаточного условия (ответ, в котором указывается лишь необходимое условие, может быть расценен учителем как неполный или неточный).

#### IV. ЕСТЕСТВЕННЫЙ ЯЗЫК И МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РЕШЕНИЯ ЗАДАЧИ

Учитывая все те выводы, которые мы сделали в связи с языковыми особенностями математических текстов, рекомендуем вам всегда проводить сначала хотя бы беглый лингвистический анализ того математического текста, с которым вы имеете дело. В качестве основных моментов такого анализа выделим:

- 1) выявление математических конструкций и соответствующих им математических символов;
- 2) поиск кванторных смыслов, как явно выраженных, так и скрытых, восполнение отсутствующих кванторных слов;
- 3) установление смысловых и логических отношений между частями условия, выявление в тексте «логических» слов.

Проиллюстрируем на крайне простом примере, что мы имеем в виду. В одном из сборников арифметических задач (для младших школьников) находим следующую задачу: *Завод за 22 рабочих дня в месяц должен был выпустить 2376 деталей. Завод увеличил ежедневный выпуск на 24 детали. На сколько дней завод ускорил выпуск намеченного количества деталей?* Чтобы понять смысл условия этой задачи и поставленного в ней вопроса, требуется (сознательно или подсознательно) определить по меньшей мере следующее:

- 1) Слово *должен* в сочетании *должен выпустить* здесь указывает не на необходимое (и не на необходимое и достаточное) условие, как в примерах, разобранных выше, а



определяет намеченное заводом 'количество деталей, которое необходимо выпустить за месяц' (по условию это число равно 2365).

2) Слово *месяц* составляет 22 рабочих дня, или срок.

3) Выражение *увеличил на* является арифметическим и соответствует сложению, причем второе слагаемое (24) дано в условии, а первое еще надо найти. В связи с этим выражением обратим внимание на то, что ребята из начальных классов часто сбиваются, пытаясь запомнить, что означают выражения *увеличить на*, *увеличить в*, *уменьшить на* и т.п. Ведь слово *увеличить* само по себе имеет лишь значение 'стать больше', а различить сложение и умножение удастся лишь благодаря предлогам. И хотя слово *увеличить* в модели может соответствовать операции умножения, в сочетании *умножить на* слово *умножить* нельзя заменить на *увеличить* (не изменяя одновременно предлога).

4) В сочетании *на сколько дней ускорил выпуск* слово *дней* следует понимать как 'рабочих дней'.

5) *На сколько дней ускорил выпуск* можно узнать с помощью вычитания из числа 22, на что указывает слово *ускорил*.

6) Слово *ежедневный* прячет в себе кванторный смысл 'каждый': *ежедневный выпуск* означает 'количество деталей, выпускаемое каждый день'.

Как мы видим, осуществляя сознательно или подсознательно данные операции, мы глубоко проникаем в смысл задачи и приближаемся к построению ее математической модели.

Вообще многие преподаватели математики не раз отмечали, что наиболее трудные вопросы на экзамене — это вопросы языкового характера. Сошлемся на сравнительно недавно выпущенную издательством МГТУ им. Н.Э.Баумана книгу доцента этого университета Э.П.Казанджана «Школьник — абитуриент — студент — инженер». В частности, в ней отмечается, что самые страшные вопросы для абитуриента: можно ли в предложении переставить, выкинуть или добавить те или иные слова, изменится ли смысл, если одни слова заменить на другие, и т.п. Далее автор пишет: «Подобные вопросы хорошо бы хоть правильно понять, где уж там правильно на них ответить...» и цитирует воспоминания академика А.Н.Крылова, где речь, в частности, идет об экзаменах во французской школе в конце прошлого века. По словам А.Н.Крылова, оценка письменных работ по математике производилась, во-первых, преподавателем французского языка, который обращал внимание на правильность языка и орфографии, а во-вторых, педагогом-математиком,

оценивавшим само содержание работы. Следует согласиться с Э.П.Казанджаном, что пристальное внимание к языку при разборе и оценке математических утверждений — вещь разумная и полезная.

Все сказанное позволяет заключить, что языковой анализ математического текста — это не выдумка или прихоть лингвиста, а необходимая и реально осуществляемая деятельность человека, работающего с этим текстом. С помощью такого анализа, как нам кажется, в ряде случаев можно не только существенно облегчить себе решение задач или доказательство теорем, но и объяснить другому ход своих мыслей (сказанное вовсе не означает, что каждый раз следует проводить анализ в явном виде). Это особенно важно, когда мы решаем задачи, математически гораздо более сложные, чем те простые примеры, на материале которых мы постарались изложить основные приемы лингвистического анализа математического текста и попутно дать некоторые практические советы.

## ЗАКЛЮЧЕНИЕ

Итак, лингвистика и математика... Надеемся, что читатель обнаружил много точек взаимодействия этих, казалось бы, далеких друг от друга дисциплин. Он познакомился с новой, сравнительно молодой областью знания, рожденной на пересечении гуманитарных и естественных наук, — формальной лингвистикой, увидел, какую пользу могут принести лингвистике точные методы анализа. Наконец, он соприкоснулся с понятиями формальной лингвистики, основными методами и приемами, увидел, какую помощь может оказать лингвисту знакомство с вычислительной техникой, с теорией и практикой алгоритмического анализа текста. Нам очень хотелось, чтобы читатель понял важность строгих и формальных рассуждений, научился отличать формальное от неформального, и прежде всего в гуманитарной сфере. И тут снова на помощь приходит математика, которая здесь уже выступает не как самостоятельная наука, не как объект изучения, а как язык, специально построенный для уточнения полуформальных и совсем не формальных утверждений. Заметим, кстати, что умением логически правильно излагать свои мысли, не путая причин со следствиями, отличая принципиально недоказуемое от доказуемого, не опуская в рассуждении логических кругов, некорректных умозаключений, полезно овладеть не только филологам.

Весьма дисциплинирует мышление искусство составления алгоритмов. Ведь без точной формулировки запросов, анализа исходных данных, умения соотнести законы естественного языка с возможностями компьютера нельзя произвести какой-либо автоматической обработки текстов на естественном языке с большим массивом исходных данных.

Читатель, наверное, понял, что математические методы применяются во всех областях языкознания: в фонетике и морфологии, морфемике и словообразовании, синтаксисе и стилистике и др. Вне нашего рассмотрения остались такие обширные области, как историческое языкознание, этимология, лингвистика текста, вопросы соотношения языка и общества, языка и мышления, проблемы понимания текста. Не следует думать, что математика обошла эти дисциплины своим вниманием. В этой связи мы можем отослать читателя к книгам Ю.Д.Апресяна и Л.П.Крысина, указанным в списке литературы. Там же можно найти дополнительные сведения, касающиеся и тех проблем, о которых шла речь выше.

По мере развития лингвистической науки математиче-

ские методы находят в ней все более широкое применение. В языкознании уже выделились целые области, где формальные приемы становятся центральными. Мы имеем в виду такие направления, как логический анализ языка, теория синтаксических структур (см. об этом в книге А.В.Гладкого), количественная лингвистика, т. е. наука, изучающая количественные соотношения в языке (см. об этом в книге Ю.Д.Апресяна), теория референции — наука, изучающая способы соотнесения языковых единиц с объектами внешнего мира (см. о ней в статье М.А.Кронгауза), структурная типология — дисциплина, изучающая сходства и различия структур разных языков. Ученые все шире используют точные методы лингвистического анализа в прикладных дисциплинах, связанных с применением различных видов электронных устройств. Особенно эффективно использование компьютеров при изучении иностранных языков, при составлении словарей, в диагностике речевых нарушений.

Лингвистическое обеспечение на равных правах с математическим формирует еще одну рожденную временем науку — информатику. Создание информационных систем хранения, обработки и выдачи информации возможно лишь на базе мощных интеллектуальных систем понимания смысла сообщений, формального представления и записи в машине исходных сведений на некотором языке. Здесь, как нигде, необходимо согласование точных математических методов с не менее точными лингвистическими.

Примечательно, что не только математика оказывает влияние на лингвистику, но и наоборот, последняя может диктовать математике свои условия, заставляя ее приспосабливаться к своим требованиям. Лингвистика требует от математики, чтобы та расширила свой понятийный аппарат, увеличила число логических и вычислительных средств. В настоящее время создаются особые языки программирования и даже целые математические системы, специально приспособленные для решения чисто лингвистических задач. Да и сам язык математики как фрагмент естественного языка постепенно становится объектом внимательного изучения со стороны лингвистов. Иногда лингвисты обнаруживают в языке математики свойства, ранее не замеченные самими математиками. Лингвистический анализ может помочь в установлении смысловой и логической эквивалентности отдельных утверждений, уточнении спорных формулировок и даже доказательства некоторых положений.

Всех, кого интересуют сложные проблемы взаимосвязи и взаимопомощи лингвистики и математики, мы можем отослать к следующей литературе:

*Апресян Ю.Д.* Идеи и методы современной структурной лингвистики. — М., 1966.

*Варга Б., Димень Ю., Лопариц Э.* Язык. Музыка. Математика. — М., 1981.

*Вентцель А.Д.* Аффинная геометрия и лингвистические задачи // Исследования по речевой информации. — М., 1968.

*Гладкий А.В.* Синтаксические структуры естественного языка в автоматизированных системах общения. — М., 1985.

*Гладкий А.В., Мельчук И.А.* Элементы математической лингвистики. — М., 1969.

*Зализняк А.А.* Лингвистические задачи // Исследование по структурной типологии. — М., 1973.

*Иванов Вяч. Вс.* Чет и нечет: Асимметрия мозга и знаковых систем. — М., 1978.

*Кондратов А.М.* Машинный перевод. — М., 1967.

*Крейдлин Г.Е., Шмелев А.Д.* Языковая деятельность и решение задач // Математика в школе. — 1989. — № 3.

*Кронгауз М.А.* Двойное пространство слова // Знание — сила. — М., 1987. — № 3.

*Крысин Л.П.* Жизнь слова. — М., 1980.

*Левин Ю.И.* Математика и языкознание. — М., 1964.

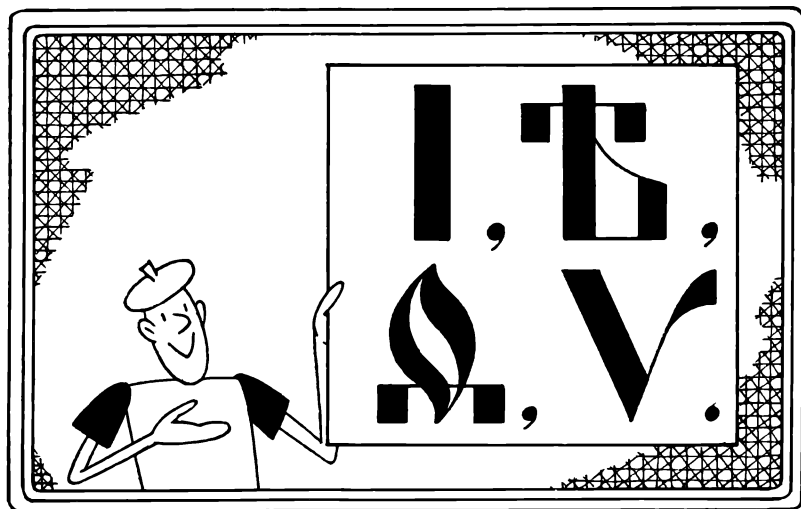
Энциклопедический словарь юного филолога. — М., 1984.

*Яковлев Н.Ф.* Математическая формула построения алфавита (опыт практического приложения лингвистической теории) // Реформатский А.А. Из истории отечественной фонологии. — М., 1970.



## ОТВЕТЫ, УКАЗАНИЯ И РЕШЕНИЯ

Приводимые в книге вопросы и задачи, естественно, различаются по сложности. Кроме того, некоторые из них носят теоретический характер. К таковым относятся задачи на построение определений важных теоретических понятий, задачи, решение которых помогает обнаружить какие-либо языковые законы, задачи, открывающие путь к построению алгоритмов. Практические задачи призваны дать читателю возможность проверить, насколько он понял текст, овладел ли он приемами построения лингвистических алгоритмов, умеет ли он осуществлять самостоятельный поиск лингвистических явлений по методике, предлагаемой в соответствующей главе. Ряд задач носит занимательный характер. Их цель — продемонстрировать существование лингвистических «неожиданностей» и парадоксов. Исходя из такого широкого спектра вопросов и задач, мы и строим данный раздел. Если задача сложная, как правило, дается ее полное решение. Кроме того, полные решения приводятся для некоторых теоретических задач. К остальным задачам даны указания и ответы.



1. Два.

2. Один.

3. Например: барабан, гиппопотам.

4\*. Очевидный и первый приходящий на ум ответ: 33 (столько букв содержит современный русский алфавит). Между тем русские тексты до 1918 г. писались по старой орфографии, в частности с использованием букв ѣ (и десятиричное), Ъ (ять), ѳ (фита), ѱ (ижица), и поэтому (с учетом и таких текстов) ответ должен был бы быть: 37. Мы видим, что ответ существенно зависит от того, как мы понимаем область определения функции  $\text{alph}(x)$ . Именно уточнение разного рода формулировок и определений и является одной из задач формальной лингвистики.

5. Верно. Взяв произвольный элемент одного множества, показать, что он принадлежит другому множеству.

6. Воспользоваться определением подмножества.

7. Поскольку в соответствии с правилами русского произношения буква  $\delta$  на конце слова обозначает глухой звук [т], утверждение опровергается.

8. а); г); е).

9. Множество знаков препинания.

10. {2; 3; 4}.

11. {И.Е. Тамм, И.М.Франк, П.А.Черенков (1958), Л.Д.Ландау (1962), Н.Г.Басов, А.М. Прохоров (1964), А.Д.Сахаров (Нобелевская премия мира 1975), П.Л.Капица (1978)}.

12\*. Для того чтобы можно было участников олимпиады рассадить по аудиториям в соответствии с указанными условиями, необходимо и достаточно, чтобы отношение «содержать хотя бы одну общую букву» было отношением эквивалентности на множестве участников. Однако в общем случае это не так. Например, пусть в олимпиаде участвуют три человека: Иванов, Барсуков, Белых. Легко увидеть, что на множестве этих трех человек наше отношение не обладает свойством транзитивности. Таким образом, утверждение, что участников в с е г д а можно рассадить так, как требуется в условии задачи, опровергается.

13. Проверить последовательно, одно за другим выполнение свойств рефлексивности, симметричности и транзитивности для каждого из этих двух отношений.

В силу условия данной задачи, а также теоремы о разбиении множество знаков разбивается на классы, такие, что в каждый класс входят те и только те знаки, которые имеют одинаковые означаемые (означающие).

14. Одна из классификаций (принятая в обычной общеобразовательной школе) может выглядеть следующим образом. Многоугольники делятся на выпуклые и невыпуклые.

Выпуклые многоугольники делятся на треугольники, четырехугольники, пятиугольники,...,  $n$ -угольники и т. д.\*. Четырехугольники делятся на параллелограммы и непараллелограммы.

До сих пор мы строили д р е в о в и д н у ю к л а с с и ф и к а ц и ю. Дифференциальные признаки легко извлекаются из определений соответствующих фигур. Далее удобно перейти (собственно, так и предлагают делать учебники геометрии) к у н и в е р с а л ь н о й к л а с с и ф и к а ц и и, используя признаки «равенство сторон» и «наличие прямого угла». Ромб — это параллелограмм, все стороны которого равны; прямоугольник — это параллелограмм с прямым углом; квадрат — это ромб, являющийся прямоугольником (или прямоугольник, являющийся ромбом). Параллелограмм, не являющийся ни ромбом, ни прямоугольником, специального названия не имеет (в таких случаях иногда говорят о «собственно параллелограммах»).

В традиционном школьном курсе геометрии эта классификация дальнейшего развития не получает. Однако бывает, что уже на первых шагах берутся другие классификационные признаки: «возможность/невозможность вписать в выпуклый многоугольник окружность», «возможность/невозможность описать вокруг выпуклого многоугольника окружность» (многоугольники, обладающие положительным значением этих признаков, называются соответственно вписанными и описанными).

Существуют и другие классификации: деление многоугольников на правильные и неправильные, на звездчатые и незвездчатые; иногда проводят деление внутри группы пятиугольников, группы шестиугольников и т. д.

15. Сможет для слов *конь*, *луч*, *стол*, *год*, *люк*, *край*, *сдал*, *пух*. Объясним, почему он не сможет этого сделать для слов *сел*, *шкура*, *шел*, а именно — почему для этих слов иностранец не сможет однозначно определить истинное звучание этих слов на русском языке.

Начнем со слова *сел*. Точки в букве ё обычно опускаются. Поэтому, видя букву *е*, иностранец не сможет решить, читается ли она как [о] или как [э].

Следующее слово *шкура*. Оно единственное в этом списке двухсложное. В двухсложных словах для их правильного чтения нужно знать, где поставить ударение, т.е. место ударения. Здесь это место однозначно не указано.

Наконец, слово *шел*. С ним ситуация такая же, как со словом *сел*.

---

\* Невыпуклые многоугольники в школе не классифицируются.





Чтобы прочитать остальные слова, требуется знать лишь общие правила чтения типа «буква *ю* после согласного читается как [у], а сам согласный смягчается» (это правило используется при чтении слова *люк*), «буква *ь* не читается, но предшествующий согласный (если это не шипящий) звучит мягко» (благодаря этому правилу удастся прочесть слово *конь*) и т. д. Все подобные правила нашли отражение в приводимом в книге алгоритме чтения, т. е. перевода буквенной записи в фонетическую транскрипцию.

16. Сможет для слов *конь*, *сел*, *стол*, *край*, *пух* и не сможет для слов *луч* (теоретически допустимо «лучь»), *год* (могло бы быть «гот»), *шкура* (на конце вместо *а* могло бы быть *о*, как в слове *скоро*), *люк* (так же читалась бы запись «люг»), *шел* (могло бы быть «шол»), *сдал* (если не знать, что *с* — приставка, можно написать «здал» — сравни *здесь*, *здание*).

17. Начальный звук может быть передан на письме единственным способом: *р*. Первый гласный звук можно записать двумя способами: *а* и *о* (сравни *просчет*); в таких случаях неединственности говорят о слабой орфографической позиции гласного и прибегают к правилам орфографии. Таким образом, первые два звука могут быть записаны двумя способами: *ра* и *ро*. Третий звук [ш'] в русских словах передается следующими буквами и сочетаниями букв: буквой *щ* (как в слове *пища*); сочетаниями двух букв *сч* (как в самом слове *расчет*), *зч* (как в слове *извозчик*), *жч* (как в словах *мужчина*, *перебежчик*), *сщ* (как в словах *расщелина*, *расщепить*); трехбуквенными сочетаниями: *ссч* (*рассчитать*), *стч*

(жестче), здч (звездчатый). Итого восемь способов. Несложный подсчет показывает, что первые три звука могут быть записаны  $2 \cdot 8 = 16$  способами.

Ударный гласный [о] после шипящего можно записать двумя способами: *е* (ё) и *о* (сравни *расческа* и *девчонка*, *дешевый* и *холщовый*). Итого для первых четырех звуков имеем 32 способа.

Наконец, согласный звук [т] в позиции конца слова записывается четырьмя способами: *т* (*счет*), *д* (*лед*), *тт* (*ватт*), *дт* (*Кронштадт*, *форштадт*). Итого общее число теоретически возможных (по правилам графики) способов написания слова [раш'от] равно  $32 \cdot 4 = 128$ . Но лишь один из них признается правильным в соответствии с существующими орфографическими правилами, а именно — *расчет*.

18. Определим сначала, что значит упорядоченность по алфавиту двух слов, т. е. что значит утверждение «слово  $\alpha$  идет по алфавиту раньше, чем слово  $\beta$ » (условное обозначение  $\alpha << \beta$ ). Напомним, что графическая запись слова — это кортеж букв. Пусть  $\alpha = \langle a_1, \dots, a_m \rangle$ ;  $\beta = \langle b_1, \dots, b_n \rangle$ . Тогда  $\alpha << \beta$ , если выполняется одно из следующих условий: 1) буква  $a_1$  предшествует в алфавите букве  $b_1$ ; 2) для некоторого натурального  $k < a_1; \dots; a_k \rangle = \langle b_1; \dots; b_k \rangle$ , а буква  $a_{k+1}$  предшествует в алфавите букве  $b_{k+1}$ ; 3)  $\langle a_1; \dots; a_m \rangle = \langle b_1; \dots; b_m \rangle$  и  $m < n$ .

Говоря неформально, сравниваются первые несовпадающие буквы слов  $\alpha$  и  $\beta$  (это формализуется пунктами 1 и 2 определения); если же несовпадающих букв нет, то раньше идет более короткое слово (это формализовано в пункте 3).

Нетрудно показать, что отношение « $<<$ » транзитивно (убедитесь в этом сами).

Список слов называется упорядоченным по алфавиту, если в нем упорядочены по алфавиту два рядом стоящих слова, различающихся написанием; в случае если два слова имеют одинаковое написание (такие слова называются омографами), то они могут в этом списке идти в любом порядке.

19. Обычный алфавитный порядок — *актер, армия, асфальт, базальт, вольт, поэт, щука*; обратный алфавитный порядок — *щука, актер, базальт, асфальт, вольт, поэт, армия*.

20. Начните с выделения классификационных признаков.

21. {Сильный ветер; шквал; резкий ветер}; {крошечная тьма; полная тьма}; крепко спать; {орать; сильно кричать; кричать что есть мочи}; вызывать слезы; связать свитер.

22. Кудахтать; {лаять; гавкать}; звонить; а) кошка; б) змея.

23. а) Вооружен; б) черный.

24. Данные сочетания представляют собою реализацию

лингвистических функций: а) Fin 'перестать быть'; Ligu 'сделать так, чтобы не (был)'.

25. *Грубая ошибка; упорное сопротивление; глубокое уважение; богатый опыт; коренное изменение; яростный гнев; тяжелая болезнь; суровое наказание; острая зависть; горячий прием.*

26. Приведем примеры еще трех нулевых морфов. Рассмотрим морфологическую форму *стен* — родительный падеж множественного числа от слова *стена*. Носителем падежно-числового значения здесь выступает нулевое окончание (сравни формы *стены, стене, стенами* и т.д.). Таким образом, здесь мы имеем нулевой морф со значением 'родительный падеж множественного числа'.

Не следует думать, что нулевые морфы встречаются только в формах существительных. Возьмем, например, форму прошедшего времени глагола *писать*. В мужском роде единственного числа она имеет вид *писал*, а в женском роде того же числа — *писала*, во множественном числе — *писали*. На основании сопоставления последних трех форм выделяется нулевой морф в форме *писал* со значением 'мужской род единственного числа'.

А вот еще нулевой морф, но уже в форме повелительного наклонения глагола. Если обратиться к форме *сядь* и сопоставить с ней формы *сяду, сядешь, сядем*, легко видеть: значение повелительного наклонения здесь можно приписать лишь знаку Ø.

27.1) Стандартная подстановка фонетического нуля на место *ы* (сравни *срыв — сорвать, подзывать — звать, засыпать — спать*); 2) стандартная замена [л] на [л'] (сравни *шла — шли, стол — столик*); 3) стандартная подстановка *ш* на место *с* (сравни *вынос — выношу, покос — кошу, воскрес — воскрешение*).

28. Отношение чередования рефлексивно по определению. Поскольку отношение чередования определяется через операцию чередования, которая не меняет означаемого морфа, то очевидно, что отношение чередования и симметрично, и транзитивно.

29. Образцом морфемы более чем с тремя представителями может служить *ХОД-*, представленная, например, следующими морфами (в буквенной записи): *ход-* (в слове *ход*), *хожд-* (в слове *хождение*), *хаж-* (в слове *хаживать*), *хож-* (в слове *хожу*).

30. Проведите доказательство от противного, воспользовавшись определением морфемы.

31. а) С приставки: *подход, выучить*.

б) С корня: *домик, писать*.

в) С суффикса — не может (доказательство с очевидностью следует из определения).

г) С окончания, вообще говоря, могло бы (ничто в определении этому не мешает), но в русском языке такого не бывает. (В некоторых языках, например в суахили, это нормальное явление. Язык *суахили* принадлежит к группе языков банту. Это язык межплеменного общения народов Восточной Африки. Суахили — государственный язык Танзании (наряду с английским), широко распространен также в Кении, Уганде, Заире.)

32. а) Корнем: *вчера*.

б) Суффиксом: *сдуру, какой-либо*.

в) Окончанием: *дом, стена*.

г) Приставкой — не может (доказательство очевидно).

33. Число корней, приставок и суффиксов теоретически неограниченно, а практически определяется объемом того словаря, которым человек пользуется. У слова, как правило, не выделяют более одного окончания. (Окончание принято считать неделимым и рассматривать его как единую морфему.) Несколько окончаний может быть лишь у слов со сложной структурой. Так, у дробных числительных (*три пятых, четыре целых сто двадцать одна тысячная* и т.п.) можно усматривать несколько окончаний, отличающихся падежными значениями; например, в форме *три пятых* окончанию *-и* приписывается значение именительного падежа (ср. *трех, трем* и т. д.), а окончанию *-ых* — значение родительного падежа.

34. Из одного (*обувь, шла*), из двух (*одеть,...*), из трех (*пар,...*), из четырех (*утюг,...*), из пяти (*старт,...*),... из девяти (*гипотенуза,...*)...

35. Вместо ответа на этот вопрос предлагаем задачу: придумать слово, корень которого содержит как можно больше звуков.

36. *Пере + под + готов + к + а; в + пуст + ую, при + зна + ний + э, при + зна + тельн + ого, чит + а + тель + ск + ий, про + дум + а + вш + ий*.

37. Один из многих возможных примеров: *знать — признать — признание*.

38. Постройте словообразовательную модель, начиная с производного слова *старый*. *Устарелость* — производное IV степени.

39. *Фронт — фронтовой — фронтовик; готовый — готовить — приготовить — подготовительный*.

40. 1) 'совершать действие, свойственное...' (*озорничать, лодырничать, хамить, рыбачить*); 2) 'становиться...' (*глохнуть, свирепеть, толстеть*); 3) 'испытывать...' (*голодать,*

нуждаться, горевать, блаженствовать); 4) 'действовать с помощью...' (пилить, бороновать, бинтовать, сверлить); 5) 'становиться как...' (звереть, каменеть, столбенеть).

41. Укажем словообразовательные средства, на основании которых выделяются соответствующие словообразовательные типы: 1) *а* (озорничать), *нича* (лодырничать), *и* (хамить, рыbachить); 2) *ну* (глохнуть), *е* (свирепеть, толстеть); 3) *а* (голодать), *а...ся* (нуждаться), *ова/ева* (блаженствовать, горевать); 4) *и* (пилить, сверлить), *ова* (бороновать, бинтовать); 5) *е* (звереть, каменеть); *ене* (столбенеть).

Некоторые из приведенных словообразовательных средств омонимичны. Например, *а* как формальное средство участвует в совершенно различных словообразовательных типах: *а* в слове *озорничать* имеет значение 'совершать действие, свойственное...', а в слове *голодать* — 'испытывать'.

Дополнительное задание. Приведите для каждой категории (из пяти вышеприведенных) еще по одному-два словообразовательных типа, входящих в эту категорию; для каждого из типов укажите соответствующее словообразовательное средство.

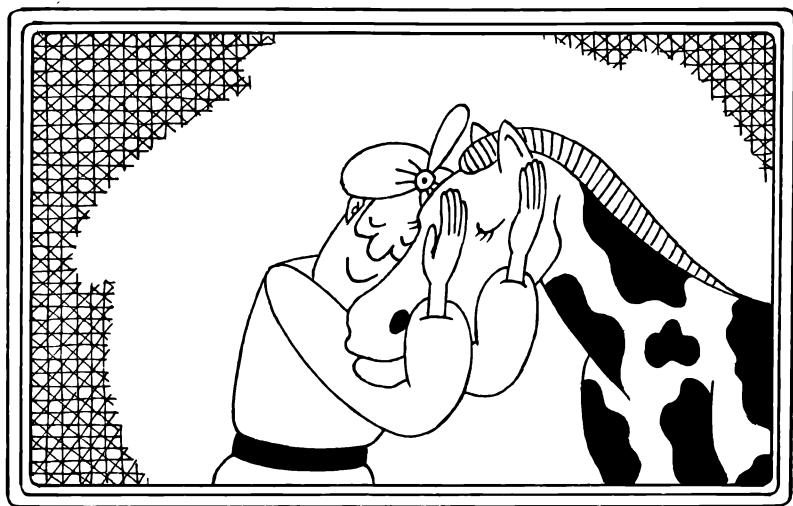
42. 1) *Полтора, два, оба* (все эти слова изменяются по родам и падежам); 2) *сотня* (изменяется по числам и падежам); 3) *сто, кто, несколько* (изменяются только по падежам); 4) *сотый, какой, двадцать первый* (изменяются по родам, числам, падежам).

43. Укажем еще по одному слову из каждого разряда: 1) *двадцать два*; 2) *стол*; 3) *пять*; 4) *маленький*.

Возможно, некоторые из вас предложили в качестве примеров слов, изменяющихся только по падежам, существительные, имеющие исключительно формы единственного числа или исключительно формы множественного числа (такие слова, как *молоко, медь, лай, сани, чернила*). С такими примерами можно было бы согласиться, если бы не одна сложность. В лингвистике существует разное отношение к этим словам. Часто таким словам, как *медь* или *лай*, приписывают словоизменительную категорию числа, но говорят, что формы множественного числа у них «неупотребительны» или «малоупотребительны», а у слов же типа *сани* иногда выделяют «формы единственного числа, омонимичные формам множественного числа».

44. Один из возможных ответов: *Один шах очень любил своего коня; Я владею тайной обучения животных человеческому языку.*

45. Для того чтобы показать, что деревья этих предложений непроективны, следует нарисовать расположенное дерево подчинения каждого из предложений и найти, где происходит



пересечение стрелки и перпендикуляра. Хочется обратить ваше внимание на четвертый пример. Формально можно было бы приписать ему проективное дерево, отражающее как цельное сочетание *отечеством свободы просвещенной*. Между тем смысл этих поэтических строк иной: слово *свободы* подчинено не слову *отечеством*, а слову *заря*. Пушкин ставит здесь вопрос, взойдет ли над отечеством заря просвещенной свободы. Именно дерево, выражающее «пушкинский» смысл, будет непроективным.

46. От древесного изображения к скобочной записи.

1) Ищем терминальные узлы, в которые идут стрелки из одного и того же узла; 2) соответствующие этим терминальным узлам слова обрамляем скобками; 3) из дерева вычеркиваем данные терминальные узлы вместе с входящими в них стрелками (тем самым в новом дереве терминальным будет узел, из которого выходили стрелки); 4) эта процедура повторяется до тех пор, пока не останется один узел («вершина»).

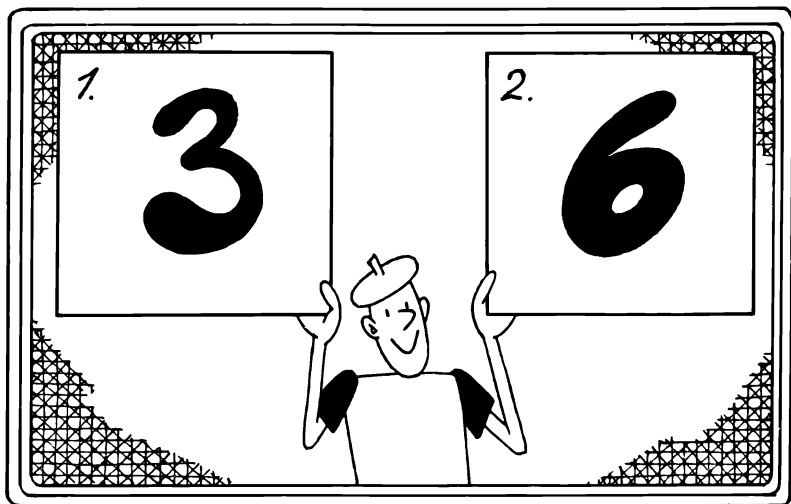
От скобочной записи к древесному изображению. Каждому слову ставится в соответствие терминальный узел, получающий изображение в виде точки. 1) Идем слева направо по скобочной записи до первой закрывающей скобки, а от нее возвращаемся назад до ближайшей открывающей скобки. Мы получили фрагмент скобочной записи, внутри себя не содержащий скобок. В нем содержатся составляющие, в узлы которых пока не входят стрелки (доказательство очевидно). 2) Над узлами, поставленными в соответствие таким составляющим, изображается еще один (новый) узел,

от которого проводится стрелка к каждому из этих узлов.  
3) Скобки стираются. 4) Процедура повторяется начиная с пункта 1).

47. 1) (магазин (принимает (от населения) (старые журналы) (и книги)))); 2) (магазин (принимает (от населения) (старые (журналы (и книги)))))). Первой скобочной структуре отвечает смысл 'книги и старые журналы', а второй — 'старые книги и старые журналы'. Древесное изображение этих скобочных структур читатель без труда восстановит самостоятельно.

48. Воспользуйтесь определениями транзитивности и отношения обрамления стрелок.

49. 1)



2) Алгоритм строится в непосредственном соответствии с правилами подсчета глубины дерева.

## ОГЛАВЛЕНИЕ

От авторов . . . . .	3
<b>Глава первая</b>	
<b>Существует ли язык математики?</b> . . . . .	5
I. Язык и языки . . . . .	6
1. Подъязыки . . . . .	—
2. Язык математики как подъязык естественного языка . . . . .	7
3. Понятие метаязыка . . . . .	9
II. Математический и алгоритмический метаязыки . . . . .	10
1. Словарь . . . . .	—
2. Грамматика рассматриваемых метаязыков . . . . .	25
<b>Глава вторая</b>	
<b>Звуки и буквы. Формальный аспект.</b> . . . . .	27
I. Фонетика и графика . . . . .	28
1. Общее понятие о классификации . . . . .	—
2. Классификация звуков . . . . .	33
3. Фонетическая транскрипция . . . . .	35
II. Графика и орфография . . . . .	38
1. Проблема перехода от звучащей речи к письменной . . . . .	—
2. Алгоритмизация графических и орфографических правил . . . . .	39
<b>Глава третья</b>	
<b>Математические методы в лексикологии и лексикографии</b> . . . . .	41
I. Словари и их формальное строение . . . . .	42
1. Виды словарей . . . . .	—
2. Требования к внутреннему устройству лингвистических словарей . . . . .	46
3. Практические требования к словарю как объекту использования . . . . .	51
II. Лексические функции . . . . .	54
1. Числовые и лексические функции . . . . .	—
2. Особенности лексических функций . . . . .	56
3. Возможности использования лексических функций . . . . .	57
III. Лингвистические модели и словари нового типа . . . . .	58
1. Модель и моделирование . . . . .	—
2. Словари нового типа . . . . .	61
<b>Глава четвертая</b>	
<b>Формальное строение слова</b> . . . . .	65
I. Формальный язык морфемики . . . . .	66
1. Формальные и неформальные определения . . . . .	—
2. Основания системы формальных определений . . . . .	68



3. Формальная система определений основных понятий морфемики . . . . .	70
4. Морфологические отношения и операции . . . . .	73
5. Внутреннее строение и работа алгоритма морфемного анализа слова . . . . .	80
II. Понятия и модели словообразования . . . . .	85
1. Отношение производности. Базовое и производное слова . . . . .	—
2. Словообразовательная цепочка . . . . .	86
3. Словообразовательное гнездо . . . . .	87
4. Словообразовательная категория . . . . .	—
<b>Глава пятая</b>	
<b>Грамматические категории и их моделирование . . . . .</b>	<b>93</b>
I. Грамматическая информация о слове . . . . .	94
1. Правила поведения слов . . . . .	—
2. Понятие грамматических категорий . . . . .	95
II. Внутренняя структура грамматической категории: логическая конструкция падежа . . . . .	96
1. Модель падежа существительных в русском языке . . . . .	—
2. Проблема количества падежей у существительного в русском языке . . . . .	100
III. Грамматические категории и согласование . . . . .	101
1. Влияние категории рода на формы окружающих слов . . . . .	—
2. Согласовательные классы . . . . .	102
<b>Глава шестая</b>	
<b>Синтаксические законы . . . . .</b>	<b>105</b>
I. Формальные синтаксические понятия . . . . .	106
1. Типы синтаксических правил . . . . .	—
2. Понятия синтаксической единицы и синтаксической связи . . . . .	107
3. Синтаксические структуры и их изображения . . . . .	—
4. Типы синтаксических деревьев . . . . .	108
II. Принцип синтаксической простоты . . . . .	118
1. Простые и сложные фразы . . . . .	—
2. Гипотеза Ингве . . . . .	120
3. Глубина дерева, по Ингве . . . . .	123
<b>Глава седьмая</b>	
<b>Математика и анализ художественного текста . . . . .</b>	<b>125</b>
I. Стилистическая диагностика . . . . .	126
1. Индивидуальный синтаксис писателя . . . . .	—
2. Перевод особенностей авторского стиля на формальный язык синтаксических деревьев . . . . .	127
II. Устройство текста волшебной сказки как разновидности текста . . . . .	131
1. Понятие сюжетной формулы . . . . .	—
2. Общая характеристика сюжетных формул русской волшебной сказки . . . . .	132

III. Математические утверждения в художественных текстах . . .	134
1. Поэты о математике . . . . .	—
2. Математические образы в поэзии . . . . .	137
3. Поэтическое обыгрывание математических понятий . . .	139
4. Логические рассуждения в художественных текстах . . . .	141

## **Глава восьмая**

<b>Языковой анализ математических текстов . . . . .</b>	<b>145</b>
---	------------

I. Языковой подход к решению задач . . . . .	146
--	-----

II. Арифметические элементы в математическом тексте . . . . .	147
---	-----

1. Примеры арифметических элементов . . . . .	—
---	---

2. Математическая и языковая модель решения задач . . . .	148
---	-----

III. Логические элементы в математическом тексте . . . . .	150
--	-----

1. Кванторы и кванторные выражения . . . . .	—
--	---

2. Необходимые и достаточные условия . . . . .	154
--	-----

IV. Естественный язык и математическая модель решения задачи . . . . .	157
---	-----

<b>Заключение . . . . .</b>	<b>160</b>
-----------------------------	------------

<b>Ответы, указания и решения . . . . .</b>	<b>163</b>
---	------------

У ч е б н о е   и з д а н и е

**Крейдлин Григорий Ефимович**  
**Шмелев Алексей Дмитриевич**

## **МАТЕМАТИКА ПОМОГАЕТ ЛИНГВИСТИКЕ**

Зав. редакцией *В.Л. Склярова*  
Редактор *Г.Н. Владимирская*  
Мл. редактор *Т.Л. Тихомирова*  
Художественный редактор *И.В. Короткова*  
Технический редактор *Т.Е. Хотюн*  
Оператор *Т.В. Василькова*  
Корректор *Л.С. Вайтман*

ИБ № 14506

Набор и верстка выполнены в издательстве «Просвещение»  
на компьютерной технике с использованием редакционно-издательской  
системы *Wave4™ Bestinfo, Inc.*

Гарнитуры из библиотеки цифровых шрифтов *ParaType™*.  
Диапозитивы изготовлены в издательстве «Просвещение»  
на фото пленке совместного производства А.О. Фототех  
и издательства «Просвещение»

Лицензия ЛР № 010001 от 10.10.91. Подписано к печати 3.10.94.

Формат 84×108 1/32. Бумага типографская № 2.

Гарнитура Таймс. Печать высокая. Усл. печ. л. 9,24. Усл. кр.-отт. 10,08.

Уч.-изд. л. 10,10+0,17 обл. Тираж 12 000 экз. Заказ № 1167.

Ордена Трудового Красного Знамени издательство «Просвещение»  
Комитета Российской Федерации по печати.  
127521, Москва, 3-й проезд Марьиной роши, 41.

Областная Орден «Знак почета» типография им. Смирнова,  
214000, Смоленск, проспект им. Ю. Гагарина, 2.

