

БИБЛИОТЕКА

ЗНАНИЕ

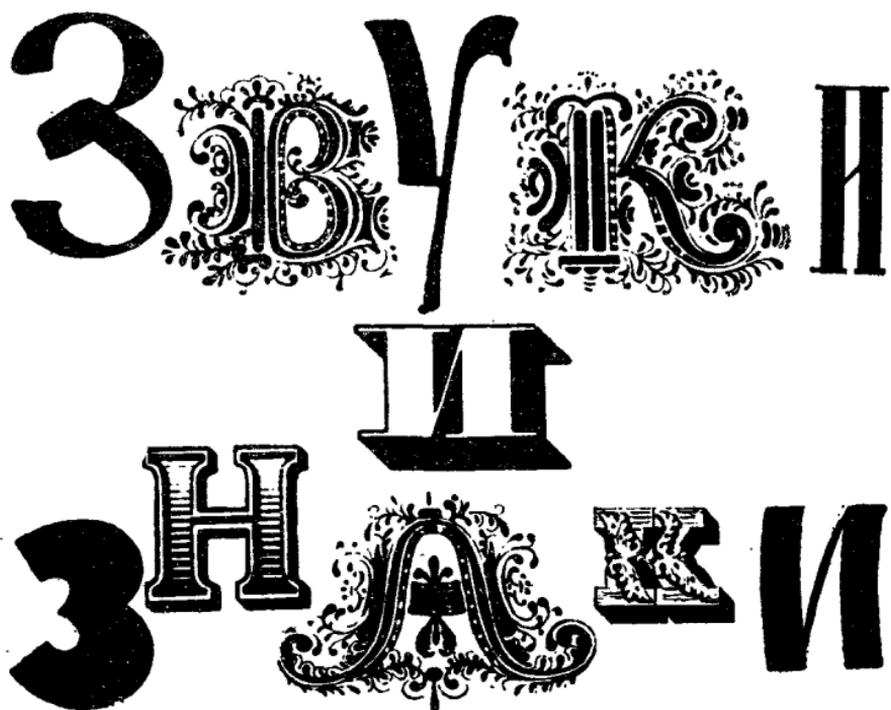
А. КОНДРАТОВ

Звуки  
и  
знаки



БИБЛИОТЕКА „ЗНАНИЕ“

*А. КОНДРАТОВ*



ИЗДАНИЕ 2-е, ПЕРЕРАБОТАННОЕ

ИЗДАТЕЛЬСТВО «ЗНАНИЕ» МОСКВА 197В

**Кондратов А. М.**

**К64** Звуки и знаки. Изд. 2-е, перераб. М., «Знание», 1978.

208 с. (Б-ка «Знание»).

Язык — это, по слову Маркса, «действительность мысли» и, как всякая другая действительность, обладает неисчерпаемым богатством содержания. Отсюда — огромное число языковедческих дисциплин, и среди них те, что родились недавно на «стыке» языкознания с математикой, кибернетикой, семиотикой (наукой о знаковых системах). Вот об этих недавно родившихся языковедческих дисциплинах, о том, чем они занимаются и к каким приходят результатам, и рассказывается в этой книге, первое издание которой вышло в 1966 г.

Автор книги — кандидат филологических наук.  
Для широкого круга читателей.

К  $\frac{70100-078}{073(02)-78}$  БЗ100-20-77

81  
4

*Моей первой учительнице,  
преподавательнице Каменской  
базовой школы Зинаиде Анфи-  
мовне Данченко эту книгу о  
языке посвящаю.*

## ДВЕНАДЦАТЬ ЛЕТ СПУСТЯ (Вместо предисловия)

В начале шестидесятых годов А. М. Кондратов, тогда начинающий автор и молодой исследователь в области кибернетической лингвистики, опубликовал свои очерки о современном языкознании и его первых контактах с математикой и кибернетикой. В 1966 году издательство «Знание» выпустило его книгу «Звуки и знаки». В нее вошли переработанные очерки, что печатались в журнале «Знание — сила», и новые, написанные специально для этой книги.

Книга «Звуки и знаки» не залежалась на полках книжных магазинов. Она была переведена на английский, французский, итальянский, эстонский и арабский языки. Ее с интересом читали не только профессионалы-лингвисты, литературоведы, историки и археологи. Эта книга стала популярной и среди представителей точных наук — математиков, кибернетиков, программистов. Вместе с тем «Звуки и знаки» использовались в качестве учебного пособия по курсу языкознания в десятках университетов и педвузов нашей страны, а отрывок из книги попал в «Учебник русского языка» для четвертого класса.

Конечно, популярность книги во многом определялась ее темой — нашим человеческим языком, неисчерпаемым хранилищем духовных богатств человеческого разума, универсальным средством общения, принадлежащим всем нам, — языком, всю сложность и богатство которого раскрывает лингвистическая наука. Однако основным секретом успеха «Звуков и знаков» состоял в том, что их автор сумел в простой и увлекательной форме профессионально рассказать о тех сложных и подчас загадочных проблемах языка, которые волновали языковедов,

математиков и кибернетиков в начале шестидесятых годов.

Теперь выходит второе издание «Звуков и знаков». Оно сохраняет ту же простоту и увлекательность изложения, сочетающуюся с высоким профессионализмом. Остался в основном неизменным круг тем, о которых она повествует.

Однако следует иметь в виду, что прошедшее десятилетие стало периодом небывало быстрого развития языкознания. Был, в частности, реализован экспериментальный, а затем промышленный машинный перевод. Начаты работы по созданию искусственного интеллекта. Была раскрыта информационная структура текста. Эти работы показали, что наш язык имеет особое нечетко-логическое построение, принципиально отличное от строго логической структуры математического языка.

Открытие «размытой» логики естественного языка, которая является более гибкой и в некоторых ситуациях более сильной, чем жесткая логика математики, заставило ученых отказаться от таких романтических проектов, как создание всеобщего языка-посредника или универсального семантического кода, с помощью которых можно было бы описать любую бытовую или научно-техническую ситуацию. Возникли принципиально новые подходы к решению таких научно-технических проблем, как машинный перевод, автоматический информационный поиск, диалог «человек — компьютер».

Для того, чтобы сделать второе издание книги злободневным, актуальным и полезным для широкого круга читателей, автору нужно было учесть те изменения, которые претерпела в общем ходе научно-технической революции наука о языке. Иными словами, из второго издания «Звуков и знаков» нужно было убрать устаревшие и поэтому малополезные знания и включить в него те новые знания, которые необходимы не только в сегодняшней практике, но и для правильного прогнозирования развития науки о языке, кибернетики, информатики и других смежных дисциплин. Автор успешно справился с этой задачей, принципиально обновив теоретическую основу своей книги.

Вот несколько примеров. Двенадцать лет назад лингвисты-структуралисты, моделируя язык, с гордостью заявляли, что они умеют совершенно формально описать не только звуковой и грамматический строй языка,

но и его словарный состав. Значение же их не интересовало — разве можно в формулах и числах выразить неуловимый и многозначный смысл слов и фраз? Если же говорить о современном языкознании, то один из его главных девизов — это «штурм значения».

Анализом значения занимаются в наши дни самые различные дисциплины, родившиеся на стыке науки о языке с другими областями знания: инженерная лингвистика и теория информации, семиотика и психолингвистика, структурная семантика и теория нечетких множеств, информатика и теория порождающих грамматик. Эта современная стратегия языкознания отражена в новом очерке «В поисках значения», включенном во второе издание книги. Он рассказывает о том, как современная наука пытается описать неуловимое значение с помощью теории нечетких множеств и вероятностной логики, толерантных пространств и статистики, психологических ассоциаций и методики «измерения значений». По существу же, этим поискам значения посвящены и все остальные очерки «Звуков и знаков».

Звуки речи, «атомы языка» — фонемы служат нам для различения смысла слов. И методика описания, созданная фонологией, начинает применяться в других областях языкознания, имеющих дело с грамматикой и смыслом, значениями слов (очерк «Формулы фонемы»). Не решив проблемы значения, невозможно развивать машинный перевод и создавать достаточно эффективные информационные языки (очерк «МП, ЯП, ИЯ»). Поискам значения все чаще уделяет внимание инженерная лингвистика, опирающаяся на статистическое описание языка (очерк «Жар холодных числ»). Язык как уникальный нежесткий «код», служащий для передачи смысла, изучает теория информации (очерк «Наш удивительный код»). Именно значение позволяет человеческому языку быть не только средством передачи информации об окружающем мире, но и своеобразным средством моделирования этого мира (очерк «Модель мира»).

Первое издание «Звуков и знаков» завершалось очерком «Лингвистика космоса». В начале шестидесятых годов, после первых космических полетов, человечество крайне оптимистично смотрело на возможность скорого общения с нашими звездными «братьями по разуму». Создавались даже специальные языки для подобного разговора, например линкос...

В наши дни, как считают крупные специалисты по космической радиосвязи, шансы человечества вступить в разговор с разумными инопланетянами весьма и весьма скромны. Вот почему второе издание книги завершается иным очерком, названным «Черный ящик». Этим термином кибернетики и психологи называют систему, устройство которой неизвестно, и мы можем судить о нем лишь по поведению этой системы.

Именно таким «черным ящиком» является человеческий мозг, который представляет собой самую сложную систему из всех, что нам известны во Вселенной. И система эта пользуется языком: сложным и вместе с тем доступным каждому человеку, произвольным и тем не менее осмысленным, нечетким и одновременно точным, универсальным и в то же самое время различным у отдельных племен и народов.

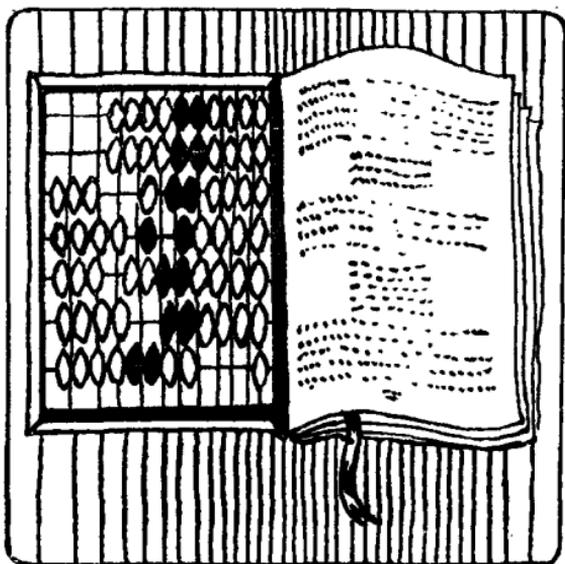
Ни одно из тех разнообразнейших средств связи, которые существуют в человеческом обществе, не может сравниться с нашим обычным языком, всеобщим богатством и достоянием. Это — подлинное чудо, принадлежащее людям, независимо от их нации, расы, образования и возраста. С помощью языка происходит превращение младенца из простого биологического существа в социального индивидуума — Человека.

Шаг за шагом раскрывает наука тайны языка. И чем больше мы их узнаем, тем больше растет наше восхищение перед этим языком. Он поистине неисчерпаем, как неисчерпаемы чудеса мира, который нас окружает и который мы не можем познать без помощи того же языка.

профессор Р. Г. Пиотровский,  
руководитель лаборатории инженерной лингвистики  
Ленинградского педагогического института  
им. А. И. Герцена

*«Числа не управляют миром, но показывают, как управляется мир», — писал великий поэт и мыслитель Гёте. В нашем веке числа начинают служить не только инженерам и физикам, но и психологам, социологам, биологам. Счет и число пришли и в науку о языке. Вот почему наш первый очерк получил название*

## ЖАР ХОЛОДНЫХ ЧИСЛ...



### О ЧЕМ ГОВОРIT СЕМИОТИКА?

«Симпозиум по структурному изучению знаковых систем» — название довольно-таки сухое. Симпозиум этот состоялся в Москве в декабре 1962 года. И всякому, кто посетил бы это собрание ученых или просто прочитал тезисы докладов, представленных на нем, стало бы ясно, что за академически официальным наименованием скрываются вопросы, затрагивающие самые различные аспекты жизни человеческого общества.

На симпозиуме читались доклады о числовом языке-посреднике и о структуре литовской народной баллады; о гадании на игральных картах как особой знаковой системе и о языковых контактах; об алгоритмах дешифровки и о пространственно-временном единстве живописного произведения; о возможностях построения структурной поэтики и о знаках рекламы, выкриках

разносчиков товаров и газет; об анализе абхазских народных игр и о транскрипции санскрита китайскими иероглифами; о функциях сложносокращенных слов и о культовом поведении в тантризме, одной из самых сложных и эзотерических, закрытых для непосвященных, систем йоги; о «тайных языках», включая воровские жаргоны, и о регулировании уличного движения как специфической системе знаков...

Один из крупнейших советских ученых не без юмора заметил, что симпозиуму недостает лишь двух докладов: первый — о моде как системе знаков, второй — о самом симпозиуме как знаковой системе. Но в словах его была не только ирония — в них содержалось и предсказание будущего.

Подавляющее число участников симпозиума, включая автора этих строк, в ту пору были совсем еще молодыми исследователями. Им казалось, что с помощью семиотики, то есть теории знаков, можно быстро и окончательно решить проблемы, над которыми вот уже сколько веков бьется человечество. Это была неизбежная дань юношескому азарту, заставляющему штурмовать «с налета» неприступные твердыни науки.

Но, с другой стороны, благодаря симпозиуму — первому научному совещанию по семиотике не только в нашей стране, но и во всем мире — развитию этой науки был дан мощный толчок. Были проведены международные конгрессы по теории знаков в Варшаве. В Советском Союзе начались интересные и обстоятельные исследования знаковых систем. Участие в них приняли видные советские философы, логики, языковеды, математики. Да и сами участники симпозиума, в ту пору еще начинающие исследователи, за истекшие полтора десятка лет превратились в ученых с мировым именем. Достаточно назвать докторов филологических наук И. И. Ревзина, Б. А. Успенского, А. А. Зализняка, М. В. Софронова и Вяч. В. Иванова, главного инициатора симпозиума по семиотике в Москве.

Теорией знаков в наши дни занимаются специалисты в самых различных областях знания. Более того, на стыке семиотики и других наук возникли самостоятельные области исследования. Назовем, например, биосемиотику, изучающую сигнализацию в животном мире с позиций теории знаков; этносемиотику, исследующую знаковые системы человеческого общества, «смысл и роль ко-

торых самими членами общества не создается»; абстрактную семиотику, родившуюся на стыке математики, логики и теории знаков; кибернетическую семиотику, рассматривающую человеческий мозг как «черный ящик», производящий операции со знаками. В Советском Союзе за последнее десятилетие вышли замечательные работы, посвященные семиотическому анализу искусства, будь то пословицы или живописные произведения, поэтическое творчество или «язык кино». Вопросы лингвистической семиотики освещались в монографиях профессоров Ю. С. Степанова, В. М. Солнцева, Ю. В. Рождественского, А. Г. Волкова, И. И. Ревзина и в работах многих других советских исследователей.

«О чем говорит семиотика» — так назывался очерк, открывавший первое издание «Звуков и знаков». Вы, вероятно, и сами догадались, почему такого очерка во втором издании книги нет. Рассказ о семиотике наших дней, о ее достижениях, проблемах, перспективах потребовал бы большой книги. Будем надеяться, что такая книга появится, ибо интерес к семиотике с каждым годом возрастает. А в этой книге, посвященной не семиотике, а лингвистике, мы коснемся лишь одного частного, но крайне важного для нас вопроса: что такое язык с точки зрения теории знаков? Что отличает его от других знаковых систем людей — таких, как жесты, этикет, дорожная сигнализация или языки программирования кибернетических машин?

Наш язык совершенно справедливо называют самой полной, уникальной и незаменимой системой связи. «Другие, искусственно созданные человеком системы и языки (например, письмо, сигнализация флажками, азбука Морзе, азбука Брайля для слепых, искусственные языки типа эсперанто или волапюк, информационно-логические языки и др.) воплощают лишь некоторые из свойств естественного языка, — пишет профессор Ю. С. Степанов. — Эти системы могут значительно усиливать язык и превосходить его в каком-либо одном или нескольких отношениях, но одновременно уступать ему в других, точно так же, как телефон, телевидение, радио (вообще всякое орудие, всякий инструмент) усиливают некоторые свойства отдельных органов человека».

Почему же наш язык, такой, казалось бы, обычный и привычный, является одновременно и самой полной, и самой совершенной, и самой богатой, и самой эконом-

ной системой знаков изо всех, что мы знаем в человеческом обществе и в сообществах животных?

Потому, отвечает семиотика, что он иерархичен. У всех остальных знаков есть выражение и содержание, означающее и означаемое. Язык же устроен гораздо сложнее.

Поясним на таком примере. Знак немислим без системы знаков. Однако один и тот же символ «!» может иметь пять совершенно различных значений. Для школьника это, вне всякого сомнения, восклицательный знак. Для шахматиста — обозначение сильного хода. Для математика — факториал. Для водителя — знак «Осторожно!». А для лингвиста — условное обозначение характерного шелкающего звука, который имеется в некоторых языках Южной Африки!

Но во всех этих случаях знак соотнесен с каким-либо понятием, звуком, пормами пунктуации. Короче говоря, это знаки, имеющие значение, заданное системой знаков... А в нашем человеческом языке?

Строго говоря, в языке знаки — это только слова. Звуки и буквы, очевидно, никакого значения не имеют. Это не знаки, а только составные части, своего рода кирпичики, или, как говорят в семиотике, фигуры, из которых строится знак.

Мы сказали выше, что знаками в нашем языке могут быть названы только слова (хотя есть точка зрения, что и слова — это не знаки, а лишь элементы знаковой системы, человеческого языка). Слова сочетаются в предложения, число которых практически бесконечно. Предложения, в свою очередь, являются элементами, из которых строится наша речь. Таким образом, перед нами иерархическая лестница: звук — корень слова или служебная частица — слово — предложение — речь или письменный текст. Причем во многих случаях один и тот же элемент языка может выступать в этой иерархии на разных уровнях.

Приведем классический пример. Два римлянина спорили, кто скажет самую короткую речь или напишет самую короткую фразу.

— *Eo gus* (еду в деревню), — таков был текст первого.

— *I*, — отвечал второй (в переводе с латыни значит: езжай!).

Рассмотрим это, действительно, предельно краткое высказывание. Во-первых, оно и в самом деле высказы-

вание, текст, речь. Состоит этот текст из одного предложения. Предложение, в свою очередь, состоит из одного слова. Слово — из одного корня. Наконец, и корень выражен с помощью одного звука или же передающей этот звук на письме буквы. Мы имеем здесь и фонетику, и морфологию, и лексику, и синтаксис! И все это — в одном значке, вертикальной палочке «I», передающей звучание «и».

Число звуков речи в любом языке мира меньше сотни. Даже в самом бедном словами языке число слов равно нескольким тысячам. Число предложений, которые можно построить, пользуясь словами, достигает астрономических величин. Число различных текстов, которые можно записать с помощью предложений, практически бесконечно. «Таким образом, язык организован так, что с помощью горстки фигур и благодаря их все новым и новым расположениям может быть построен легион знаков,— пишет датский ученый Луи Ельмслев, перекинувший мост между лингвистикой, наукой о языке, и семиотикой, наукой о знаках.— По своей цели языки прежде всего знаковые системы, но по своей внутренней структуре они прежде всего иное, а именно — системы фигур, которые могут быть использованы для построения знаков».

Благодаря этому свойству языка мы получаем возможность в любое время высказать любую мысль, любое чувство, любую фантазию или причуду. С помощью горстки фигур строится все бесконечное многообразие и богатство языка, подобно тому, как с помощью горстки химических элементов строится тот удивительный мир, в котором мы живем.

## СТРУКТУРЫ И ЧИСЛА

Анализ языка с позиций семиотики, по сути дела, открыл науке язык как таковой. До того времени лингвисты интересовались не языком как таковым, не системой знаков и составляющих эти знаки фигур, а речью, текстами, порожденными системой. Наиболее четко это выразил основоположник современной структурной лингвистики Фердинанд де Соссюр.

«Единственным и истинным объектом лингвистики является язык, рассматриваемый в самом себе и для себя»,— такими словами завершал свой знаменитый

«Курс общей лингвистики» Соссюр. Не следует думать, что это было открытым провозглашением формализма, принципиальным отказом понимать общественные функции языка и т. д. Напротив, сам Соссюр считается создателем так называемой социологической школы в языкознании, исходящей из того, что наш язык — продукт общественный, и понять его невозможно без связи с другими общественными науками и явлениями. Но есть, так сказать, лингвистика внутренняя и лингвистика внешняя, лингвистика языка и лингвистика речи.

«Наше определение языка предполагает устранение из понятия «язык» всего того, что чуждо его организму, его системе,— одним словом, всего того, что известно под названием «внешней лингвистики», хотя эта лингвистика и занимается очень важными предметами и хотя именно ее главным образом имеют в виду, когда приступают к изучению речевой деятельности,— писал Соссюр.— Язык есть система, которая подчиняется лишь собственному порядку. Уяснению этого может помочь сравнение с игрой в шахматы, где довольно легко отличить, что является внешним, что внутренним. То, что игра пришла в Европу из Персии, есть факт внешнего порядка; напротив, внутренним является все то, что касается системы и правил игры. Если я фигуры из дерева заменяю фигурами из слоновой кости, то такая замена будет безразлична для системы; но если я уменьшу или увеличу количество фигур, такая перемена глубоко затронет «грамматику» игры».

Продолжая аналогию Соссюра, мы можем заметить, что фигуры вообще могут отсутствовать: вспомним об игре опытных шахматистов вслепую, не глядя на доску, не притрагиваясь к шахматам. Если правила игры в шахматы являются системой, языком, то любая партия, которую мы станем играть, будет текстом, порожденным этой системой, «речью».

Число шахматных фигур невелико, так же как и число полей доски и число правил, которым подчиняется игра в шахматы. А вот число различных партий в шахматы настолько велико, что его можно считать бесконечным. Так и с помощью языка, состоящего из конечного числа элементов и грамматических правил, можно образовывать любое число фраз и текстов.

Язык и речь, система и текст — их взаимоотношение интересует не только лингвистику, но и семиотику. Ме-

тоды теории знаков применяются ныне в фольклористике и литературоведении, в теории музыки и театра, искусствоведении, этнографии, поэтике, инженерной и социальной психологии. И везде мы имеем дело с ситуацией *язык — речь, система — текст*, будь то «язык» жестов или «текст» человеческого поведения.

Более того, в естественных науках, вроде этологии, изучающей поведение животных, или генетики, исследующей механизм наследственности, мы также сталкиваемся с проблемой *система — текст*. Наконец и в технических дисциплинах, например в машиностроении или электронике, стоят сходные вопросы: построение из системы — конечного числа деталей или полупроводников — неисчерпаемого количества «текстов» машин, транзисторов и т. п.

Современную структурную лингвистику сравнивают иногда с логиком для целого ряда общественных и естественных наук. Именно она первой осознала знаковый характер языка и отличие его от речи. Систему языка структурные лингвисты стараются описать в строгих терминах математической логики и семиотики. Естественно, что тексты описаны так быть не могут. Ведь число их практически бесконечно!

Здесь вместо формул и символов нужны иные методы — и в первую очередь математической статистики, тот «жар холодных чисел», о котором писал Блок в своих «Скифах», позволяющий, казалось бы, сухим колонкам цифр превращаться в яркие картины, показывающие скрытые механизмы языка, порождающего речь. Самое же ценное в этом знании — это то, что оно начинает в наши дни все чаще оказывать помощь людям в их практике, в повседневной жизни. Например, в обучении языку.

### **ЧИСЛА, УЧЕБНИКИ И ЭВМ**

Ни для кого не является секретом, что есть учебники хорошие и есть плохие. Опытный педагог порекомендует вам такой-то учебник, а такой-то братъ отсоветует... Но на каком основании? На основании опыта, интуиции, наконец, своих взглядов. Но взгляды, как известно, бывают разные. Вот если бы можно было вычислять степень эффективности того или иного учебника! Тогда

вместо споров о его достоинствах и недостатках можно было оперировать точными числами.

На первый взгляд все это кажется фантастикой. Однако первые попытки в этом направлении уже делаются. Не так давно в Минске с помощью ЭВМ сравнивалась эффективность учебников иностранных языков, и эта разработка белорусских ученых была удостоена серебряной медали ВДНХ... Вот вам и фантастика!

В основе знания языка — знание слов. Но сколько именно этих слов нужно дать в одном уроке учебника? И сколько слов вообще должен содержать словарный минимум? Чтобы новые слова усвоились, они должны встречаться и далее в учебнике, это очевидно. Но каким должен быть интервал между повторениями одного и того же слова?

Долгое время все эти вопросы решались на глазок, у одного педагога своя точка зрения, у второго — своя, третий считает ошибочными обе эти точки зрения и предлагает свою собственную. В наши дни союз педагогов, лингвистов, психологов, программистов и ЭВМ позволяет решать подобного рода вопросы с помощью точных чисел. Например, установлено, что темп введения новых слов не должен превосходить пяти процентов от всех слов, что встречаются в учебном тексте данного урока. А текст, содержащий 3,6 процента новых слов, оптимален — и прежние слова хорошо повторяются, и новые лучше всего усваиваются.

Чем большее число раз встретилось слово в уроке и во всем учебнике, чем более равномерно распределено оно по всему учебнику, тем учебник лучше. Интуитивно мы все понимаем это. Но как доказать интуицию числами? Попробуйте-ка произвести нужные подсчеты — и вы убедитесь, что без вычислительной машины тут никак не обойтись. Нужно подсчитать, сколько раз встретилось то или иное слово в каждом уроке и в упражнениях, сопровождающих урок. Нужно высчитать средние арифметические и средние квадратические отклонения для каждого слова по каждому уроку, затем суммировать данные по всему учебнику.

В одном учебнике ряд слов будет иметь лучшие показатели, чем те же показатели в другом. С другой стороны, в «худшем» учебнике будут и слова, и целые уроки с лучшими показателями, чем в первом. Параметров же,

по которым вычисляем мы все эти показатели, у нас несколько. Значит, нужно еще вычислять по формулам значимость того или иного параметра... Словом, без ЭВМ во всей этой бухгалтерии не обойтись.

Вот такие подсчеты и провели белорусские исследователи, поручив ЭВМ сравнивать эффективность учебников. Разумеется, это только начало кибернетизации педагогики (ведь дело не только в количественных данных лексики, но и в ее качестве, да и не только в лексике). Трудно дать сразу ответы на все вопросы, над которыми бьется педагогика вот уже столько веков.

Еще одним конкретным примером союза лингвистики, математики, педагогики и ЭВМ являются работы, проводимые во многих странах мира и связанные со словами и числами.

### СКОЛЬКО НУЖНО СЛОВ?

В любом языке мира насчитываются тысячи, десятки тысяч, а то и сотни тысяч различных слов, как, например, в русском, английском, немецком, французском и других, имеющих давнюю письменную традицию. Все слова не знает даже носитель языка (вряд ли кто из русских знает значение тех четырехсот сорока тысяч слов, что собраны в академической картотеке Института русского языка).

Значит, изучая иностранный язык, мы должны заучить не все, а самые нужные, самые важные слова. Именно их должен включать словарь-минимум, что обязательно прилагается почти к каждому учебнику языка. Но тут возникает вопрос: сколько же слов должно быть в этом словаре-минимуме?

Одни специалисты полагают, что для чтения учебного текста достаточно и семисот пятидесяти слов, а при запасе в две с половиной тысячи можно читать простой неадаптированный текст с помощью словаря. Другие называют тысячу восемьсот — столько слов, по их мнению, достаточно, чтобы понимать речь на чужом языке и вести на нем повседневное общение.

Шестая Международная ассоциация преподавателей современных языков рекомендует тысячу пятьсот — три тысячи слов на шесть лет обучения. Нашей программой для языковых вузов и факультетов иностранных языков установлен минимум в пять тысяч слов, причем на

первом курсе должно быть усвоено восемьсот слов. Программа для неязыковых вузов рекомендует на два года обучения минимум в две с половиной тысячи слов, из них тысячу двести — для активного усвоения.

Согласимся с программой для вузов. Однако тут встает очередной вопрос: а какие именно слова нужно включать в словарь-минимум? Понятно, что самые важные, самые употребительные. Предлоги, местоимения, союзы... Правда, их не так-то уж и много в любом языке. А дальше? Какие глаголы, прилагательные, существительные нужно знать наизусть, а какими можно и не обременять свою память?

Казалось бы, словари-минимумы в учебниках должны совпадать. Если не полностью, то хотя бы на три четверти, хотя бы наполовину... Когда же специалисты сравнили шестнадцать разных учебников французского языка для американских школ, то обнаружили, что общими для них является примерно два процента слов! В учебниках насчитывалось в общей сложности шесть тысяч разных слов. И только сто тридцать четыре из них входили во все словари-минимумы!

Представим, что два американца, изучив французский по разным учебникам, попробуют общаться друг с другом на этом языке. Вряд ли они поймут друг друга: ведь общими у них будет менее полутораста слов.

Ту же печальную картину показал анализ и других словарей-минимумов. Десять учебников испанского языка содержали четыре с половиной тысячи слов, общими же было двести сорок девять. А в двадцати шести учебниках того же испанского языка было обнаружено тринадцать тысяч слов. Хороший же словарь-минимум для тех, кто изучает испанский язык! Требовать от новичков, изучающих иностранный язык, знания всех этих слов нелепо и бессмысленно, не всякий переводчик-профессионал их знает.

Какие же слова считать употребительными, а какие нет? Ответ на этот вопрос дает статистика, так называемые частотные словари.

## ЧАСТОТЫ И СЛОВА

Чем чаще встречается в текстах слово, тем оно употребительнее, или, говоря языком математики, имеет большую частоту. Но как избежать слов, характерных

только для данного текста, а в других почти не встречающихся? Самый лучший путь — брать не один источник, а несколько. А затем, подсчитав, сколько раз встречаются в этих книгах различные слова, сравнить результаты. Если они совпадают, то слово действительно является частым (или редким). Если не совпадают, то, значит, в одном из текстов оно было завышено из-за сюжета или пристрастия автора к этому слову.

Так, весьма частое для «Капитанской дочки» слово *крепость*, встретилось всего лишь в четырнадцати источниках из ста тридцати трех. Значит, слово это отнюдь не частое и занимает скромное место в списке самых употребительных слов русского языка. А составляй мы список лишь по «Капитанской дочке», то оно вошло бы едва ли не в первую двадцатку самых ходовых слов нашего языка!

В «Истории Пугачева» оно встречается еще чаще — сто тридцать четыре раза (ведь там речь то и дело идет о крепостях, взятых Пугачевым). В «Путешествии в Арзрум» оно встречается шестнадцать раз. Зато во всех художественных произведениях (кроме «Капитанской дочки») Пушкин употребил слово *крепость* всего один раз (в «Дубровском») и два раза в стихотворениях и поэмах.

Но не только количество разных текстов может гарантировать наши подсчеты от случайности. Не менее важно взять большой объем этих текстов. И чем он больше, тем вернее мы гарантированы, что никакие случайные факторы не повлияют на точность наших подсчетов.

Подсчеты частот слов производились давно: еще в средние века составлялись индексы к религиозным текстам, в которых указывалось, сколько и где упоминаются слова *бог*, *ангел* и т. д. Но только на пороге нашего века появляется серьезная потребность в частотных словарях. Первым таким словарем считается словарь немецкого языка, выпущенный в Берлине в 1897—1898 годах. Его составитель Ф. Кединг обработал колоссальный материал в одиннадцать миллионов слов!

Вслед за тем приходит черед частотных словарей английского языка. В Лондоне в 1904 году священник Дж. Ноулз составляет список из слов, которые встретились более двадцати пяти раз в текстах, объемом в сто тысяч слов.

Американец Р. Эддридж, управляющий небольшой фабрикой, где в основном работали иммигранты, за два

года расписал на карточки двести пятьдесят статей из газет, составивших около пятидесяти тысяч слов, и подсчитал частоты этих слов.

Первоначально, как говорил Элдридж в предисловии к своему словарю, выпущенному в 1911 году, его целью было выявить минимум слов, который помог бы рабочим-иностранцам быстро овладеть разговорной речью. Но затем, продолжает Элдридж, он понял, что частотные словари могут принести и более существенную пользу, в том числе оказать помощь в реформе английской орфографии, которую справедливо называют «чудовищной».

Но частотные словари Ноулза, Элдриджа, а также ряд других, выпущенных в начале нашего века, были все-таки любительскими, ненадежными. Ноулз опирался на текст Библии, Элдридж — на текст газетных статей, другие авторы — на выборки из различных писем. Да и объем исследуемого материала был невелик. В двадцатых годах профессора Колумбийского университета Э. Торндайк и И. Лордж начинают работу над серией больших частотных словарей английского языка. Завершился их четвертьвековой труд выпуском крупнейшего в мире частотного словаря. Обработав различные тексты общим объемом в восемнадцать миллионов слов, ученые привели в своем словаре тридцать тысяч слов, которые встретились в этих текстах.

В настоящее время частотные словари созданы почти по всем европейским языкам. Словарь испанского языка, составленный Гарсиа Осом, создан на основании подсчетов четырехсот тысяч слов; словарь чешского языка — одного миллиона двухсот тысяч слов, польского — семи миллионов, французского — полутора миллионов и т. д.

Пришел черед и для частотного словаря русского языка. В настоящее время составлено пять таких словарей.

### **ПЯТЬ ЧАСТОТНЫХ СЛОВАРЕЙ**

Интерес к русскому языку возрастает во всем мире. Русский учат в США и Анголе, в Японии и на Кубе, в Финляндии и Монголии, в Канаде и Эфиопии. Сколько же слов надо знать иностранцу, чтобы читать произведения русской классики? вести деловой разговор? объясниться с прохожим, приехав гостем в СССР?

Мы уже говорили о картотеке Института русского языка, где зарегистрировано четыреста сорок тысяч слов. В семнадцатитомном «Словаре современного русского языка» приводится сто пятьдесят тысяч различных слов. Вряд ли целесообразно иностранцу, решившему постичь великий и могучий русский язык, заучивать все эти слова, которые в полном объеме неизвестны и русским людям. Ему нужен на первых порах минимум, слова самые употребительные, самые частые.

В 1953 году в американском городе Детройте выходит в свет первый частотный словарь русского языка. Его составитель Г. Г. Джоссельсон использовал тексты общим объемом в один миллион слов. Но так как основным материалом для анализа Джоссельсон взял произведения дореволюционных писателей, то в список часто употребляемых слов попало слово *барин* и другие подобные ему умершие в социальном смысле слова. Да и методика обработки материала была несовершенной. Короче говоря, словарь этот не отражал реальной картины современного русского языка и был лишь первым опытом.

Десять лет спустя в Таллине вышел «Частотный словарь современного русского литературного языка», составленный Э. А. Штейнфельдт на основании обработки четырехсот тысяч слов.

Однако и этого словаря оказалось недостаточно. Объем текстов, взятых для словаря Штейнфельдт, был невелик. Причем четвертую часть его составляли произведения детской литературы. Естественно, это повлияло и на общую картину: в список частых попали такие слова, как *пионер*, *штаб* и т. п.

В 1966 году университет штата Огайо выпустил частотный словарь «разговорного русского языка». Составлен он был на основании обработки пьес. Мы же, как известно, далеко не всегда говорим так, как персонажи пьес классических и даже современных. Вот почему университет дружбы народов имени П. Лумумбы в Москве выпустил в 1968 году свой словарь, более точно отражающий разговорную лексику. В словарь вошло две тысячи триста восемьдесят наиболее употребительных слов русской разговорной речи.

Наконец настал черед и монументального «Частотного словаря русского языка». Такой словарь был издан в нашей стране в 1977 году и занял почти тысячу

страниц. Его составителями были сотрудники филологического факультета Ленинградского университета и лаборатории семиотики НИИ прикладной математики и кибернетики при Горьковском университете. С помощью ЭВМ они обработали тексты объемом в миллион слов.

В словаре в алфавитном порядке от союза *а* до прилагательного *ящичный* приводится около сорока тысяч различных слов, встречавшихся в тексте. Рядом со словом дана его частота, сначала общая, а затем по основным четырем категориям, на которые были разделены все тексты (художественная проза; драматургия; научные и публицистические тексты; газетные и журнальные тексты). Тут же приведено и число разных текстов, в которых встречалось данное слово (ведь эта характеристика не менее важна, чем абсолютная частота, которая, как уже говорилось, может зависеть от специфики текста, сюжета и т. п.).

Вслед за алфавитно-частотным словарем идет частотный. Слова даны здесь в порядке убывания частот. На первом месте стоит предлог *в*. Вместе со своим вариантом предлогом *во* он встретился в текстах почти сорок три тысячи раз! На втором месте по частоте употребления идет союз *и*, он встретился более тридцати шести тысяч раз. Отрицание *не* — на третьем месте (около двадцати тысяч раз), на четвертом — предлог *на* (свыше семнадцати тысяч). Местоимение *я* занимает пятое место — оно встретилось в текстах около четырнадцати тысяч раз.

Верхнюю часть списка наиболее употребительных слов занимают предлоги, союзы, частицы, местоимения. Лишь в четвертом десятке этих слов появляется глагол *мочь*, числительное *один*, а на шестом десятке — существительное *год* и прилагательное *большой*. Зато потом эти части речи занимают львиную долю списка.

## ТЕКСТООБРАЗОВАТЕЛЬНАЯ СПОСОБНОСТЬ

Третья часть «Частотного словаря русского языка» посвящена статистическим данным. И статистика, приводимая в ней, нужна не только специалистам по математической лингвистике или теории вероятностей. Она имеет важное значение в практике, в отборе слов для изучения русского языка.

В текстах, объемом в миллион слов встретилось около сорока тысяч словоупотреблений. Но свыше тринадцати тысяч различных слов употреблены были всего лишь один раз, около шести тысяч — два раза. Ясно, что слова эти редкие, их нельзя включить в словари-минимумы. Столь же ясно, что слова, имеющие частоту в несколько тысяч, в эти словари должны быть включены обязательно. Легко провести черту между очень частыми и очень редкими словами. Но как провести ее между словами не очень частыми и скорее частыми, чем редкими? Сколько нужно слов для словаря-минимума?

Вопрос этот мы уже задавали в самом начале нашего рассказа. Теперь, когда вы узнали о частотных словарях, переформулируем его: какую часть слов из списка самых частых мы должны брать — сотню слов? тысячу? десять тысяч?

На помощь лингвистике вновь приходит статистика.

Обратимся к нашему «Частотному словарю русского языка». Первые десять слов (ничтожнейшая часть от общего числа различных слов) покрывают почти двадцать процентов всего текста. Первая сотня самых частых слов составляет лишь четверть процента словаря. Зато в тексте она покрывает свыше сорока процентов всех слов. Первая тысяча слов, два с половиной процента всех слов словаря, покрывает свыше шестидесяти семи процентов текста.

Около девяти тысяч слов встретилось в текстах десять и более раз. Слова эти составляют менее четверти всего словаря. Зато покрывают они более девяноста процентов текста. Остальные же тридцать тысяч слов не покрывают и восьми процентов всех текстов. Очевидно, заучивать эти тридцать тысяч слов нецелесообразно. Знание же слов с частотой «десять» и более позволит понимать практически весь текст.

Может быть, и в списке слов, имеющих частоты «десять» и выше, также можно сделать сокращения? Первая тысяча слов частотного словаря, будь то русский, испанский, английский, французский, немецкий или любой другой язык, покрывает значительную часть текста. Какую — зависит не только и не столько от конкретного языка, сколько от стиля, типа речи.

В английском языке тысяча самых употребительных слов покрывает около семидесяти процентов литератур-

ных текстов и свыше восьмидесяти процентов — научных. Во французском языке та же тысяча самых частых слов покрывает около восьмидесяти процентов литературных текстов и свыше девяноста — разговорной речи.

Но какими бы ни были вариации, ясно, что чем больше мы будем удаляться от «верхушки» частотного словаря, тем больше понадобится нам различных слов для покрытия (а стало быть, и перевода) текста.

Возьмем французский язык. Первая тысяча слов дала возможность покрыть свыше девяноста процентов разговорной речи. Но если мы добавим еще тысячу слов, то прирост будет равен всего четырем процентам. А следующие три тысячи слов увеличат покрываемость текста немногим более трех процентов.

Примерно та же картина получится, если мы возьмем другие языки или другие стили. В среднем первая тысяча самых частых слов покрывает около восьмидесяти процентов всего текста. Увеличение списка еще на одну тысячу дает прирост порядка шести процентов. Добавление следующей, третьей, тысячи позволит покрыть около трех процентов, четвертой — около двух процентов, пятой — немногим более одного процента текста.

«Пользуясь этими данными, можно разумно планировать введение лексики в ходе преподавания иностранного языка,— пишут советские специалисты, говоря о помощи, которую статистика и лингвистика оказывают преподаванию языков.— Так, если согласиться с тем, что за одно занятие учащиеся неспециального вуза могут усвоить 12—18 словоформ, то через 7—10 занятий они должны будут знать не менее 120 словоформ, покрывающих более чем 50% текста. В этом случае учащийся должен узнавать и переводить по отдельности каждое второе слово неадаптированного текста данной специальности. Через 30—40 занятий учащемуся будет знакомо уже около 500 словоформ, покрывающих в среднем около двух третей всех словоупотреблений текста, а к концу первого учебного года учащийся сможет узнавать около 1000 словоформ, покрывающих около 80% текста».

### СЛОВАРИ, СЛОВАРИ, СЛОВАРИ...

В чем отличие частотного словаря от любого другого? Прежде всего в том, что вместе со словом в нем приводится и частота употребления этого слова в тексте.

Слова могут идти либо в алфавитном порядке, либо по убыванию частот. В «Частотном словаре русского языка», как мы уже говорили, дается и алфавитный, и частотный списки. Частотный словарь может быть полным, когда в нем приводится список всех слов, встретившихся в тексте, и неполным, когда дается лишь список слов, встретившихся с определенной частотой. Так, алфавитный список «Частотного словаря русского языка» — полный, а частотный — неполный, он дает лишь список слов с частотой «десять» и более, остальные тридцать тысяч слов, менее употребительных, в него не включены.

Объем материала, который лег в основу словаря, в значительной мере влияет и на состав этого словаря. В зависимости от того, сколько тысяч или миллионов слов текста обследовалось — или просчитывалось на электронно-вычислительной машине — выделяют малые, средние и большие частотные словари (так, первые частотные словари английского языка были, безусловно, малыми, а словарь Торндайка и Лорджа, составленный по текстам в восемнадцать миллионов слов, относится, разумеется, к большим).

Однако на словарь влияет и стиль, и тематика, и жанр, и авторская принадлежность того или иного текста. Вот почему различают частотные словари устной и письменной речи, словари общие (вроде нашего «Частотного словаря русского языка») и отраслевые (например, Е. А. Калининой с помощью ЭВМ составлен частотный словарь русского подъязыка электроники). Есть словари отдельных произведений, как классиков, так и современных авторов. Составлены «персональные» частотные словари — Пушкина, Шекспира, Шевченко и других.

Не менее важно и то, что именно фиксируется в словаре: исходная форма слова, словоформа, то есть слово в различных грамматических формах, основа слова или словосочетание. В таких языках, как китайский, разница между словарем основ и словоформ практически неощутима. В английском языке она уже заметна. Когда же мы обращаемся к такому языку, как русский, где множество падежей, глагольных форм и т. д., различие между исходной формой слова и словоформой огромно. Например, частотный словарь русского подъязыка электроники, составленный по текстам в двести тысяч слов, дал свыше двадцати одной тысячи различных словоформ и менее семи тысяч слов!

Частоты, приводимые рядом со словами, также различны. Это может быть абсолютная частота, указывающая, сколько раз в тексте встретилось то или иное слово. Частота эта может быть относительной, выраженной в процентах или вероятностях (ведь объемы словарей могут быть разными, в зависимости от них различны и абсолютные частоты). Во многих словарях приводятся и другие числовые характеристики, например количество источников, в которых встретилось то или иное слово.

Наконец, словари могут различаться по способу анализа материала. Обычно их составляют на основании выборок. Так, для «Частотного словаря русского языка» брались отрывки из произведений различных авторов — писателей и драматургов, ученых и публицистов. В их числе были Ленин и Калинин, Леонов и Шолохов, Вавилов и Тарле, Тимирязев и Ферсман.

Однако есть и другой метод — сплошное расписывание текста. Именно так составляется словарь языка писателя. Подобного рода словари начали составляться давно. В них включаются все слова, которые употребил тот или иной автор в своих произведениях. Существует уже множество словарей, посвященных отдельным произведениям, например «Божественной комедии» Данте, «Неистовому Орланду» Торквато Тассо, «Улиссу» Джеймса Джойса и другим (кстати сказать, «Улиссу» принадлежит своеобразный рекорд по числу различных слов: если в «Божественной комедии» их пять тысяч восьмисот шестьдесят, в «Неистовом Орланде» — восемь тысяч четыреста семьдесят четыре, то Джойс ухитрился употребить почти тридцать тысяч разных слов в одном романе!)

Словарь языка писателя нетрудно сделать и частотным — для этого надо только указать, как часто употреблялось то или иное слово. Именно таким словарем является «Словарь языка Пушкина», выпущенный в нашей стране в четырех томах. В нем приводятся данные о том, сколько раз встречается то или иное слово в Полном собрании сочинений Пушкина, в каких именно произведениях, в каких значениях и грамматических формах.

Около шестисот тысяч слов содержат все пушкинские тексты. Из них двадцать одна тысяча двести слов — различные. Свыше ста раз встречается лишь семьсот двадцать слов. Зато один раз во всех пушкинских текстах

употребляется шесть тысяч четыреста сорок различных слов, два раза — две тысячи восемьсот восемьдесят слов, три раза — одна тысяча восемьсот слов. Это — неопровержимое и точное свидетельство богатства пушкинского словаря, своеобразия его лексики.

Это, так сказать, общая статистическая картина. В последнее время в нашей стране появляются работы, в которых делается попытка на основании частотного словаря дать характеристику своеобразной модели мира, которая лежит в основе творчества каждого большого поэта. Большие поэты, как известно, проходят на своем пути этапы развития, вехи которых — поэтические сборники или поэмы. Сопоставление частотных словарей, составленных по отдельным сборникам, позволяет выразить в числах и точно сформулировать то, что интуитивно чувствует читатель.

Вот, например, сравнение частотных словарей по сборнику стихов Бориса Пастернака «Сестра моя жизнь» и Осипа Мандельштама «Камень» (сборники эти вышли почти одновременно и признаны едва ли не высшими достижениями в творчестве этих поэтов). Если откинуть все служебные слова, союзы и прочий «грамматический фон», то самым частым существительным у Пастернака является *ночь*, вслед за ним идут слова *глаза*, *губы*, *звезда*, *сад*. В «Камне» Мандельштама самое частое существительное — *Рим*, затем *мир*, *сердце*, *печаль*.

Еще более разительный контраст получается, если сгруппировать существительные по «смысловым полям»: природа, вещи, человек, культура и история. Слова, относящиеся к «полям» человек и вещи, употребляются обоими поэтами примерно одинаково. Но если у Пастернака «поле» природа включает половину всех слов, то у Мандельштама — менее тридцати процентов. И, наоборот, «поле» культура и история у Пастернака занимает три процента слов, а у Мандельштама — почти двадцать!

Смысловое «поле» флора у Пастернака содержит свыше сотни названий растений и слов, имеющих прямое отношение к растительному царству (грядка, шишка и т. п.). У Мандельштама в «Камне» лишь двенадцать названий растений. Пастернаковский «зоопарк» состоит почти из полусотни названий живых существ, от бациллы и стафилококка до ехидны и мамонта. Еще

двадцать пять слов связаны с животным миром (рыба, гнездо, хвост и т. п.). Фауна Мандельштама бедней, всего лишь двадцать одно животное.

Сопоставление подобного рода смысловых «полей», частот отдельных слов и групп слов ярко показывает различие между моделью мира, запечатленной Пастернаком в сборнике «Сестра моя жизнь», где чувства человека слиты в неразрывное целое с окружающей природой, и моделью мира Мандельштама, выраженной в сборнике «Камень», где история и культура органически связаны с человеческим бытием. А если бы у нас был частотный словарь по произведениям Маяковского, написанных в ту же пору, когда писались «Камень» и «Сестра моя жизнь», мы бы наглядно убедились в том, что модель мира великого пролетарского поэта была совсем иной, она обращена была на революцию, опрокинувшую старый строй.

### СТИЛИСТИКА И СТАТИСТИКА

К сожалению, помимо многотомного «Словаря языка Пушкина» мы имеем лишь частотный словарь «Стихов о Прекрасной Даме» Блока, да два словаря, о которых только что шла речь выше. Вероятно, в недалеком будущем у исследователей будет достаточное число таких словарей. И тогда можно будет сопоставлять творчество различных поэтов одной эпохи (например, Блока и Брюсова, Маяковского и Хлебникова, Пастернака и Есенина, Мандельштама и Цветаевой, Багрицкого и Уткина), проводить сопоставление различных сборников и поэм одного и того же поэта (скажем, сопоставить «Сестру мою жизнь» не только с «Камнем» Мандельштама, но и с другими произведениями самого Пастернака — «Волнами», «Темами и вариациями», «Спекторским» и т. д.).

Не только на основании частотных словарей получаем мы возможность сделать доказательным то, что чувствуем лишь интуитивно, подсознательно. Методы статистики все шире проникают в литературоведение, поэтику, стилистику. С их помощью исследователи могут давать характеристику стилей различных авторов не только качественную, но и количественную.

Так, уже простой подсчет среднего количества слов в предложении позволяет характеризовать стиль того или

инного писателя с помощью чисел. Как пишет советский математик Р. Л. Добрушин, «можно сказать: А. Н. Толстой предпочитает более длинные фразы, а А. И. Куприн — более короткие». А можно сказать и так: «Среднее число слов в фразе в произведении Толстого «Сестры» равно 11,9, а среднее число слов в фразе в произведении Куприна «Поединок» — 9,5». Разница будет примерно такая же, как если в одном случае ограничиться утверждением, что производство угля в Советском Союзе больше, чем в Англии, а в другом — привести конкретные цифры. Каждому ясна большая доказательность утверждения во втором случае.

Первую попытку применить статистику в литературоведении сделал почетный академик Н. А. Морозов, известный деятель «Народной воли», более четверти века проведший в одиночном заключении в Шлиссельбургской крепости. Правда, методика исследования его была несовершенна, что и отмечал крупнейший русский математик А. А. Марков сразу после выхода работы Морозова «Лингвистические спектры». «Только значительное расширение поля исследования (подсчет не пяти тысяч, а сотен тысяч знаков) может придать заключениям некоторую степень основательности, если только границы итогов различных писателей окажутся резко отделенными, а не обнаружится другое весьма вероятное обстоятельство, что итоги всех писателей будут колебаться около среднего числа, подчиняясь общим законам языка», — писал Марков в «Известиях Академии наук» в 1915 году.

Действительно, подсчеты показывают, что стиль отдельных писателей можно характеризовать статистически, употребление отдельных частей речи, типов предложений и т. д. подчиняется определенной вероятностной закономерности.

Приведем несколько примеров. На пятьсот знаменательных слов у Куприна приходится семьдесят семь глаголов, у Пушкина — сто десять, у Чехова — сто двадцать семь. В прозе Симонова на пятьсот слов приходится сто семьдесят существительных, сорок девять прилагательных, семьдесят три местоимения, сто одиннадцать глаголов. У Шолохова соответственно двести шестнадцать существительных, семьдесят семь прилагательных, тридцать девять местоимений, семьдесят семь глаголов.

Еще более характерны числа, говорящие о соотношениях между частями речи у того или иного автора. «Лермонтов видит и изображает мир, действительность в большем разнообразии качественных характеристик, признаков, чем это делает Пушкин: в среднем у Лермонтова 39 имен существительных из каждых 100 получают признаки прилагательных, а у Пушкина таких существительных всего 25; значит, речь и мышление Пушкина предметнее, Лермонтова — «качественнее». О сходной особенности речи и мышления двух авторов говорит и соотношение «наречие — глагол»: в среднем у Лермонтова на 100 глаголов приходится 42 наречия, а у Пушкина — всего 26; а это означает, что речь и мышление Лермонтова активнее окрашивают воспринимаемые процессы, чем это делают речь и мышление Пушкина. Интересно и то, что соотношение «существительное — глагол» и «глагол — существительное и прилагательное» оказываются у Пушкина и Лермонтова статистически равными, что позволяет опровергать гипотезу об особой глагольности речевого стиля Пушкина», — пишет Б. Н. Головин в книге «Язык и статистика».

А вот какие интересные данные были получены после подсчета названий цветов у разных авторов в их прозе и стихах. В стихах об Америке Маяковского на тысячу слов текста приходится восемь цветообозначений, а в прозе, говорящей о той же Америке, — только два слова на тысячу. В военной лирике Симонова названия цветов употребляются в восемь раз чаще, чем в его же рассказах военных лет.

### ВЕЗДЕСУЩИЕ ЧИСЛА

Мы уже говорили, что стиль каждого большого поэта и писателя имеет свои количественные характеристики. Они начинают служить филологам и литературоведам, позволяя решать спорные вопросы об авторстве с помощью чисел. Так, кстати сказать, решили давний вопрос «о дедушке Гомере»: был ли автором «Илиады» один человек или же, как полагали многие специалисты, «Илиада» — сборник героических песен, лишь по традиции приписываемый великому слепцу.

Текст «Илиады» набили на перфокарты, затем ЭВМ тщательно проанализировала, то есть пересчитала все

ритмические особенности каждой главы эпоса. Подсчеты машины неопровержимо показали: автором поэмы мог быть только один человек. Все главы «Илиады» сохраняют общее ритмическое единство.

С помощью чисел ученые начинают решать не только практические, прикладные задачи языкознания и литературоведения, но и многие теоретические вопросы. Например, определять степень заимствования одного языка из словарного запаса другого. Как известно, заимствуются обычно слова, обозначающие предметы быта, культуры, труда, которых не было ранее у того или иного народа (естественно, что вместе с предметом заимствуется и его название). Например, русское слово *спутник* вошло во все языки мира, равно как французское *одеколон*, арабское *жирафа*, австралийское *бумеранг* и т. п.

Все языки мира равны, на любом из них можно выразить все, что выражено на другом языке. Заимствования не говорят о том, что тот язык, из которого заимствуются слова — лучший, а который заимствует — худший. Однако разные языки по-разному восприимчивы к этим заимствованиям. Удивительной консервативностью обладает исландский язык. Иностранные слова практически в него не попадают, а новому понятию или термину подбираются свои собственные определения, средствами самого исландского языка. И вот футбол по-исландски звучит как *кнаттспурна*, то есть пинание мяча; дыня — как *троллепли* (яблоко великана), мотороллер — это *трещащая гадюка*, кинофильм — живой образ, ракета — огненный полет и т. п.

Другие языки, наоборот, весьма восприимчивы к иностранным словам (вероятно, многие читатели знают о знаменитой полемике, которую вели в прошлом веке романтики и архаисты во главе с адмиралом Шишковым, предлагавшим калоши именовать мокроступами). В албанском языке, как показывают подсчеты, из пяти тысяч ста сорока слов собственными являются лишь четыреста тридцать слов, все остальные заимствованы из других языков. В армянском языке полторы тысячи слов из одной тысячи девятисот сорока заимствованы из персидского, греческого, сирийского, парфянского, арабского языков. В корейском языке до семидесяти процентов заимствований из китайского; в современном английском от пятидесяти до семидесяти процентов всех слов заим-

ствованы из французского, латыни и другие романских языков.

Ученые смогли проследить динамику этих заимствований. Известный датский лингвист Отто Есперсен исследовал несколько томов «Большого Оксфордского словаря» английского языка, в котором собрано около полумиллиона различных слов. Вслед за ним А. С. Бо провел подсчеты по всем томам этого монументального издания. Оказалось, что если принять количество заимствований в течение 1100—1600 годов за сто процентов, то около шестидесяти процентов слов было заимствовано в 1100—1400 годах, около двадцати — в 1401—1500 годах и примерно столько же — в 1501—1600 годах.

Однако эти подсчеты говорят лишь о том, сколько слов дожило до нашего времени. Обычный словарь ничего не скажет, сколько слов французского языка не дожило, хотя они и были заимствованы в то или иное время англичанами. Тут на помощь приходят словари не простые, а частотные.

Еще в 1947 году американский ученый Дж. Ципф обнаружил любопытнейшую связь между частотой употребления слова и его «возрастом». Чем выше эта частота, тем древнее слово. И, наоборот, слова с незначительной частотой, как правило, появились в языке сравнительно недавно (вы можете легко проверить по «Частотному словарю русского языка», о котором мы рассказывали). Там, где не помогает традиционная методика подсчета заимствований, срабатывает новая методика — по частотным словарям. Вот какие цифры были получены на материале частотного словаря английского языка. Оказывается, с 1100-го по 1400 год не шестьдесят, а все девяносто процентов слов было заимствовано из романских. На другие два периода (1401—1500 и 1501—1600 годы) остается лишь по пяти процентов от общего числа заимствованных слов.

Была найдена и не менее интересная зависимость между временем появления слова в английском языке и его длиной в слогах. Почти половина всех односложных слов в английском языке имеет возраст в восемьсот и более лет. Такой же солидный возраст у двусложных слов, но уже не половины от их общего числа, а только одной пятой. Трехсложных слов этого возраста — три процента, четырехсложных — один процент. А среди пя-

тисложных слов в английском языке нет ни одного, которое бы имело возраст в восемь столетий — все они мо- ложе!

Быть может, связь между временем и словами языка (английского, русского или любого другого языка мира) выражается не только в устойчивости слов, имеющих разное число слогов, но и в изменении всего словаря? Или по крайней мере какой-то его части? Нельзя ли най- ти лингвистические часы, подобные «часам» геологи- ческим, хронологическим, астрономическим, с помощью которых мы определяем время событий?

Эта мысль пришла в голову американскому языкове- ду Морису Свадешу по аналогии с методом датирова- ния по распаду радиоактивного углерода. Лингвистичес- ким часам был посвящен целый очерк в первом издании этой книги. Их называли «почти точным инструментом», позволяющим датировать события, о которых, казалось бы, не осталось никаких вещественных памятников или памятников письма. Как же смотрит наука на эту про- блему в наши дни?

### ЛИНГВИСТИЧЕСКИЕ ЧАСЫ

Проанализировав скорость изменения нескольких языков и языковых семей, Свадеш нашел, что она равна примерно восьмидесяти двум плюс-минус два процента за тысячелетие. То есть в языке за тысячу лет сохраняется около восьмидесяти процентов слов, входящих в ядро, в основную лексику языка...

Так ли это? Свадеш проанализировал древнеегипет- ский, китайский, романские языки. Если ход лингвисти- ческих часов объективен, как часов радиоуглеродных, то, стало быть, близкие результаты можно получить и на материале любых других языков мира. Однако когда ученые стали проверять эту «среднюю константу скорос- ти», взяв другие языковые семьи, оказалось, что она не является всеобщей.

Вот несколько примеров. По мнению археологов и лингвистов, единый праславянский язык стал распадаться в конце VI века до н. э. Однако подсчеты, проведен- ные по методу Свадеша, дали совсем иную дату. Русский и чешский языки, оказывается, разошлись четы- реста-пятьсот лет назад, чешский и польский—двести-че- тыреста лет назад. Всякому, кто хоть немного знает

историю своей страны и братских славянских народов, ясно, что эти числа смехотворно малы (более тысячи лет назад чехи имели и свою государственность, и свое письмо, и свой язык!).

Поразительно неубедительную дату получили по методу лингвистических часов, когда сравнили два скандинавских языка — старонорвежский и исландский, чтобы определить время их расхождения. Известно, что к 930 году завершилось заселение Исландии, в основном выходцами из Норвегии. Между тем лингвистические часы показали не тысячу и даже не полтысячи лет, а всего шестьдесят три — сто девяносто четыре года. На самом деле уже в прошлом тысячелетии исландцы имели и свой язык и свою литературу.

Правда, неточность хода лингвистических часов в этом случае можно оправдать той уникальной консервативностью исландского языка, о которой мы уже говорили выше... Ну, а славянские языки? Или, как показали исследования востоковедов, скорость изменения армянского языка, языков Средней Азии и ряда других, не соответствующая формуле Свадеша? Под вопрос стал сам метод датирования с помощью словаря, динамики изменения слов, его составляющих. Или, быть может, дело не в самом методе, а в лексике, тех словах, которые мы берем за эталон?

Свадеш составил список из двухсот пятнадцати слов, потом сократил его до сотни. Но почему, собственно говоря, до ста? А может быть, надо расширить этот список до пятисот? Проверка показала, что разные слова имеют разную устойчивость. Русский язык — индоевропейский, числительное *три* звучит в нем так же, как звучало оно в языке древних хеттов, как звучит оно в священном языке Индии — санскрите, сходным образом звучит оно и в английском, немецком и других индоевропейских языках. А вот понятие «мальчик» меняло свой словесный ярлык по несколько раз чуть ли не в истории каждого языка. В древней Руси говорили отрок, а потомки древнерусского языка — русский, украинский и белорусский — сменили это слово. Русские вместо отрока говорят мальчик, а украинцы и белоруссы — хлопчик.

Вот почему современные лингвисты начинают проверять слова на их устойчивость в языке, степень сохранения того или иного слова в зависимости от понятия, которое оно обозначает. Более того, оказывается, что

есть связь между частотой употребления слова и временем его возникновения в языке. Не так давно в нашей стране вышла книга «Математические методы в исторической лингвистике», авторы которой М. В. Арапов и М. М. Херц предложили математическую модель лингвистических часов, где работы Сवादеша оказываются лишь частным случаем более общего метода.

По сравнению с методом Сवादеша, «шаг вперед состоит здесь в том, что нет более необходимости требовать, чтобы скорость изменения словаря всегда была постоянной и имела одинаковое значение для различных языков,— пишут Арапов и Херц.— Таким образом, вместо списка слов, выражающих фиксированные, тщательно выбранные понятия, можно было бы взять случайно выбранный фрагмент словаря и проверять, имеют ли слова из него соответствия в родственном языке».

Формулы, найденные советскими учеными, показывают, что случайная выборка из словаря распадается точно так же, как и весь словарь. А это значит, что такая случайная выборка подходит для целей датирования не меньше, чем отобранная сложным методом сотня-другая слов. Иными словами, механизм лингвистических часов оказался и более прост, и более сложен, чем это представлялось его первооткрывателю Морису Свадешу.

Впрочем, такую простоту и одновременно сложность мы находим всякий раз, когда начинаем изучать язык с помощью статистики. Числа помогают лингвистам исследовать язык во всех его измерениях, начиная со слов и кончая субъективными ассоциациями, связанными со словами. Но за числами стоят, очевидно, какие-то закономерности, определяющие статистику.

Нельзя ли увидеть за фактами формулы? Вскрыть с помощью математики механизмы языка, порождающие все многообразие нашей речи?

### ОТ ФАКТОВ К ФОРМУЛАМ

Статистические данные отражают в числах рост, убывание или стабильность различных элементов языка. Но они не вскрывают механизм процесса, его динамику. Вот почему в настоящее время языковеды, имея дело с числами, стараются строить на их основании математические модели, которые не только отражают динамику, но и позволяют делать прогнозы на будущее и «загляды-

вать» в прошлое, о котором нет достоверных данных. Вот несколько подобного рода моделей, предложенных ленинградскими лингвистами А. А. Пиотровской и Р. Г. Пиотровским.

В русских научных и электротехнических текстах XIX века слова типа *вольт*, *рентген*, *радиан* во множественном числе родительного падежа писались так: *вольтов*, *рентгенов* и т. п. Однако, как показала Л. К. Граудина, начиная с конца восьмидесятых годов прошлого века, все чаще стали употребляться написания *вольт*, *рентген*, совпадающие с именительным падежом единственного числа. Спустя два-три десятилетия эти формы утвердились не только в профессиональной речи, но и в литературном языке. В итоге появилась новая группа имен существительных, которая в родительном падеже множественного числа имеет нулевое окончание: мы говорим и пишем: тысяча вольт, пять рентген, а не вольтов или рентгенов.

Числовые данные, характеризующие динамику этого процесса, можно свести в таблицу (например, если в 1885 году написание типа *вольт* встречалось один раз на сотню, то в 1908 году — уже девяносто девять раз).

Данные таблицы были перенесены на график, где по оси абсцисс отмечались годы, а по оси ординат — частоты форм с нулевым окончанием. «Полученная последовательность экспериментальных точек показывает резкое возрастание нулевых форм в период между 1886 и 1905 гг. Возникает вопрос, какой из функций можно воспользоваться для описания полученной зависимости? — пишут Пиотровские. — Линейная зависимость здесь применена быть не может, поскольку значения функции находятся в интервале от  $-\infty$  до  $+\infty$ , в то время как по условиям задачи область изменения нашей функции лежит в интервале между нулем и единицей (относительные частоты не могут быть меньше нуля и больше единицы)». Рост нулевых форм лучше всего моделирует график обратной тригонометрической функции  $f = \arctg t$ , где  $f$  — частота нулевых форм, а  $t$  — годы.

Пример этот имеет иллюстративный характер — все числовые данные у нас были. Однако часто лингвисты имеют дело с отрывочными сведениями, неполными материалами по диалекту, эпохе или стилю того или иного языка. Здесь математическая модель помогает восстано-

вить не засвидетельствованные в дошедших до нас памятниках этапы развития языка. Так, А. А. Пиотровская и Р. Г. Пиотровский выводят формулу, по которой можно вычислить динамику формирования и развития в старофранцузском языке определенного артикля (формирование это шло в народно-разговорной речи, которая почти не отражена в дошедших до нас памятниках той эпохи).

Зависимость между объемом текста, который подвергается обработке, и числом разных слов, которые в нем окажутся, очевидна. Нельзя ли отыскать математически строгую формулу, по которой можно было бы, исходя из объема текста, вычислять количество слов? И определять, какой объем даст нам статистически достоверные результаты?

Первым найти такую формулу словаря попытался уже упоминавшийся нами Дж. Ципф. Связь между частотой употребления слова и его рангом, то есть номером в списке, получила наименование «закон Ципфа». Частотные словари представляют собой обычно списки слов, которые расположены по их рангу: первым идет слово, которое встречается чаще всего, затем второе по встречаемости и т. д. Однако выяснилось, что «закон Ципфа» не универсален. Были попытки описать распределение слов в тексте с помощью специальных формул теории вероятностей — так называемого нормального распределения, распределения Пуассона, распределения Маркова — Колмогорова и т. д. (причем, как показала советская исследовательница М. Е. Каширина, распределение Маркова — Колмогорова является наиболее общим и универсальным для распределения любых языковых единиц).

В теории вероятностей известны десятки законов распределения случайной величины. Задача статистической лингвистики — выбрать тот закон, который лучше всего отражает именно реалии языка, а не какие-либо иные закономерности.

Вот характерный пример, заимствованный нами из учебника «Математическая лингвистика», написанного Р. Г. Пиотровским, К. Б. Бектаевым и А. А. Пиотровской. И наше обычное поведение, и функционирование техники, и порождение речи — в той или иной степени вероятностны. Садясь в самолет или автомобиль, мы

уверены, что все будет хорошо. Составляя словарь для перевода русских текстов по математике, мы не станем включать в него слово *дядя* или словосочетание *бубновый туз*.

И все-таки несчастные случаи, увы, бывают, какова бы ни была их вероятность. В книгах по математике можно найти и бубнового туза и даже дядю (так, в труде «Теория вероятностей» Е. С. Вентцель читатель может обнаружить цитату из начала «Евгения Онегина», знаменитое «Мой дядя самых честных правил...»). Так что же, отменить автомобили и не летать на самолетах? А в математические словари наряду со словом *дядя* включать еще и *тетю*, и *бабушку*, и названия игральные карт и вообще все сотни тысяч русских слов? Разумеется, нет.

Если сравнить астрономически большое число полетов и автомобильных пробегов с числом несчастных случаев, станет ясно, насколько мала их вероятность. И вероятность всех этих дядей и бубновых тузов в математических трактатах мала — хотя авторы их могут и процитировать Пушкина, и воспользоваться известными всем игральными картами, иллюстрируя пример случайного выбора или комбинаторных сочетаний.

Студент сдает экзамен. Из ста предложений, данных ему для перевода, в шести он напутал с синтаксисом. Пятерки такой студент не заслужил, но зачет ему поставит любой здравомыслящий преподаватель, даже не знакомый с теорией вероятностей. Ибо понимает, что с помощью словаря такой студент сумеет перевести любой взятый наугад текст. Но если такое же число ошибок сделает машина-переводчик, ее программа зачета не получит. Студент умеет пользоваться словарем, при переводе опирается на смысл фразы, и небольшие помехи с синтаксисом ему не слишком повредят. А ЭВМ свои ошибки в синтаксисе не искупит ни лексикой, ни смыслом, ей недоступным. Прощать мы должны не шесть, а скажем, одну ошибку на сто фраз.

Что же касается техники, тут дело и вовсе серьезное. Вот почему так строго к нарушениям наше ГАИ и так тщательно проверяют готовность самолета работники Аэрофлота. Ибо тут, когда речь идет о людях, случайность должна быть сведена до минимума: не одну ошибку на сотню случаев, а даже одну ошибку на сто тысяч нельзя допускать!

Связь техники, статистики и языкознания наметилась давно. Ведь даже на простой, но очень важный вопрос: как удобнее расположить клавиши на пишущей машинке? — нельзя ответить одному только технику или лингвисту. Нужно знать частоты употребления различных букв и сочетаний этих букв. Нужно знать конструкцию машинки (вот почему редкие буквы расположены на периферии, а частые — в центре клавиатуры, причем на основании статистики спарены *t* и *b*, *n* и *p* и т. д.).

Изобретение телеграфа, телефона, других средств связи вызвало сближение инженерии, математики и лингвистики. Их союз помогает решать задачи кодирования слов и букв языка в электрические сигналы, сокращения «избыточных» частей текста при передаче телеграмм.

Однако подлинный союз техники, лингвистики и статистики в изучении языка начался с появлением электронных вычислительных машин. Если раньше необходимые подсчеты требовали многих лет труда и большого коллектива сотрудников, то теперь с помощью ЭВМ они могут быть произведены очень быстро и экономично.

Мы упоминали частотный словарь немецкого языка, составленный Кедингом на основании одиннадцати миллионов слов. Разумеется, один человек такую работу проделать не мог: не хватило бы всей его жизни. В составлении словаря Кедингу помогали сотрудники. Как вы думаете, сколько их было? Десяток, сотня, две сотни? Нет, гораздо больше — тысяча человек!

В течение нескольких лет вел кропотливую работу над составлением «Словаря языка Пушкина» коллектив сотрудников Института русского языка Академии наук СССР. Четверть века вели свою работу над частотным словарем английского языка Торндайк и Лордж...

В наши дни составление этих словарей поручено электронным вычислительным машинам. Именно они являются соавторами всех больших частотных словарей, появившихся в последние годы, включая «Частотный словарь русского языка».

С помощью ЭВМ в наши дни осуществляется издание монументального, в пятнадцати томах, словаря французского языка. Он охватывает почти шестьсот произведений авторов XVIII—XX столетий, общим объемом в семьдесят миллионов (!) слов. Каждому из семидесяти

тысяч слов, вошедших в словарь, дается толкование, приводится общая частота его употребления, а также частоты, с которыми оно употреблялось в том или ином веке. Нет сомнения в том, что и для других языков появятся столь же нужные словари. Конечно, без машины составление их вряд ли было бы возможно: вспомните Кединга и его тысячу помощников.

ЭВМ помогают и в составлении частотных словарей отдельных «подъязыков», охватывающих ту или иную область науки и техники: электронику, терапию, строительные материалы, автомобилестроение, геологию нефти и газа и т. п. В нашей стране составлено более десятка таких словарей для английского, немецкого, французского и других языков. Осуществляет это важное и жизненно необходимое для практики дело группа «Статистика речи».

Правда, объем текстов, взятых для словаря, сравнительно невелик: выборка состоит, как правило, из двухсот тысяч слов. Но на это есть свои причины. Вы помните, что львиная доля текста покрывается самыми частыми словами, первой их тысячей. Вторая тысяча частых слов даст незначительный прирост, третья — еще меньший и т. д. Чем больший объем текста мы возьмем, тем больше разных слов в нем будет, но прирост этот будет замедляться. Вот что показывают расчеты. В словаре английского языка, составленном Г. Кучерой на основе машинной обработки миллиона слов, встретилось около пятидесяти тысяч различных слов. Если увеличить выборку в десять раз, то есть до десяти миллионов, число разных слов возросло бы не в десять, а примерно в два с половиной раза. Если довести выборку до ста миллионов слов, число разных слов увеличилось бы не в сто, а лишь в четыре раза.

Еще меньший прирост дают тексты по отдельным подъязыкам и областям знания. Вот какая картина рисуется нам, если мы попробуем применить ЭВМ для составления полного частотного словаря английских текстов по электронике. Двести тысяч слов, обработанных на машине, дали около десяти тысяч различных слов. Если увеличить выборку в пять раз, до миллиона слов, то число разных слов даже не удвоится. Увеличь мы выборку в пятьдесят раз, число разных слов возрастет только в три раза; увеличение в пятьсот раз, до ста миллионов слов, даст прирост в четыре раза. И если

мы доведем выборку до фантастической величины в десять миллиардов, то есть в пятьдесят тысяч раз, то число разных слов возрастет лишь шестикратно. Наш словарь охватит лишь шестьдесят тысяч разных слов, между тем терминология электроники гораздо богаче, эти слова ее не исчерпывают, хотя мы обработали бы на ЭВМ около тридцати миллионов страниц текстов!

Вот почему ученые находят более рациональное применение вычислительной технике. С помощью машин выявляются не все слова в той или иной области науки и техники (где употребляются сотни тысяч различных терминов), а наиболее употребительные, а также самые частые словосочетания. Ведь они, например, в отраслевых англо-русских словарях, занимают почти три четверти объема словаря. Выявить же эти сочетания, одновременно подсчитывая их частоту вручную, практически невозможно: лишь ЭВМ способна проделать столь утомительную и однообразную работу.

Но не только электронные вычислительные машины оказывают помощь лингвистике. В свою очередь, помощь современного языкознания оказалась крайне необходимой для вычислительной техники. С рождением кибернетики появляется и проблема «разговора» человека и робота. ЭВМ понимают только строго формализованный, однозначный язык чисел и логических команд. Можно ли перевести на него наш обычный человеческий язык?

Машинный перевод с одного языка на другой, машинный перевод устной речи в письменную и письменной в устную, наконец, разговор с машиною «по-человечески», то есть ввод и вывод информации в машину посредством устной речи — все эти задачи несколько десятков лет назад относились к области научной фантастики. А ныне они стали реальными и важными проблемами практики. И решать эти проблемы можно лишь в тесном содружестве математиков, техников и лингвистов.

Инженерная лингвистика — так называют это направление в научной литературе. И речь здесь идет не о том, чтобы инженер обучился основам языкознания, а лингвист — математике и электронике. Инженерная лингвистика в наши дни превратилась в отрасль исследований, имеющих свои конкретные задачи, свою методику и свои объекты изучения. Причем, как отмечает руководитель группы «Статистика речи» Р. Г. Пиотров-

ский, решая собственные задачи, инженерная лингвистика стимулирует появление новых идей не только в области языкознания и смежных наук, но и в таких, казалось бы, далеких от нее дисциплинах, как методика преподавания языков, структурное литературоведение, радиолокация, психиатрия. Но особое значение для инженерной лингвистики имеет ее связь с кибернетикой и электронными вычислительными машинами.

Каждая идея и разработка инженерной лингвистики должна быть реализована на ЭВМ. Без машины, без робота теряется сама суть, смысл этого направления. Первые попытки машинного перевода состояли в том, чтобы передать машине поиск слов в словаре. Задолго до наступления эры кибернетики, в 1933 году, советскому инженеру П. П. Троянскому было выдано авторское свидетельство на изобретение «машины для подбора и печатания слов при переводе с одного языка на другой или несколько других одновременно». Но чем дальше шли работы в области машинного перевода, или, как говорят специалисты, МП, тем яснее становилось, что осуществить этот перевод можно лишь тогда, когда мы сумеем заставить машину переводить по-человечески, не по форме, а по смыслу..

Так у инженерной лингвистики появилась стержневая проблема: автоматическое распознавание смыслового образа. И проблема эта связана с другими проблемами распознавания образа машиной — зрительного для читающих автоматов, звукового для автоматов, действующих по приказу, который дает человеческий голос.

Тут от, казалось бы, сугубо технических проблем инженерная лингвистика переходит к проблемам, стоящим на повестке дня авангарда современной науки: моделированию мышления, работы человеческого мозга...

*В нашем веке сугубо теоретические и сугубо практические задачи оказываются неразрывно связанными друг с другом. И, пожалуй, наиболее ярко эта связь теории и практики проявилась в области машинного перевода и создания информационных языков.*

*Об этом расскажет очерк*

## МП, ЯП, ИЯ



### ОТ КРУГА ЛУЛЛИЯ ДО ЭВМ

Раймонд Луллий, живший в XIII столетии, был человек, далеко не заурядный. В молодости рыцарь, он оставил светскую жизнь, дабы найти эликсир бессмертия. Эликсир этот, несмотря на десятилетия алхимических поисков, найти ему не удалось. Зато, по мнению Луллия, он открыл основной компонент эликсира — воду жизни, на латыни — аква вита. Впоследствии оказалось, что этой водой жизни был чистейший спирт, который умели возгонять задолго до Луллия арабские алхимики.

Специалисты по вычислительной технике, структурные лингвисты и ученые, занимающиеся теорией знаков, все чаще начинают упоминать в своих работах другое изобретение экс-рыцаря и алхимика — его называли «кругом Луллия».

Во времена Луллия считалось, что в каждой области знаний есть небольшое число принципов, которые столь же неколебимы, как аксиомы математики: их следует принимать без доказательств. Это так называемые абсолютные начала — доброта, величина, сила, сознание, воля, правда, слава; относительные начала — различие, согласованность, противоположность, середина, конец, равенство; добродетели — справедливость, умеренность, вера, милосердие, надежда и пороки — жадность, обжорство, сладострастие, гордыня, лень, зависть, гневливость, лживость, предательство; субъекты — бог, ангел, небо, человек...

Всего у Луллия, согласно тогдашней схоластике, получилось шесть категорий. В каждой из них было ровно по девять элементов. Элементы эти Луллий разместил на шести concentрических кругах. Первый, внутренний, круг состоял из девяти основных вопросов: что? почему? из чего? сколько? где? когда? какое? которое из двух? каким образом? Теперь, если привести систему кругов во вращение, можно получить, по мысли Луллия, ответ на любой вопрос, связанный с аксиомами, начертанными на кругах. Иными словами, Луллий считал, что его вертушка охватывала все знания, которые может вместить наш разум!

Как видим, наш разум Луллию представлялся неким ограниченным вместилищем знания, а сами знания чем-то завершенным, окончательным. Правда, «емкость» этого вместилища (или хранилища) получалась огромной. Число возможных комбинаций действительно было велико: вы в этом легко убедитесь сами, если возведете число элементов — девять — в шестую степень (число категорий). А так как круги изготовлялись из металла и раскрашивались в различные цвета, то вертушка Луллия производила весьма сильное впечатление на современников. Рассказывают, что Луллий с помощью своих кругов хотел обратить в христианство владыку мусульманского города, но тот, не поддавшись магии кругов, продал изобретателя в рабство, и единоверцам Луллия пришлось платить большой выкуп...

Но так или иначе изобретение Луллия является прообразом нынешних устройств, с помощью которых человек хочет усилить мощь своего разума — логических машин. Идея формализации мышления, сведения его к вычислениям и логическим операциям высказывалась

двумя гениальными учеными XVII столетия Декартом и Лейбницем. «Подобно тому, как можно в один день научиться на каком-нибудь неизвестном языке называть и писать числа до бесконечности, таким же образом должна быть найдена возможность сконструировать все слова, необходимые для выражения всего, что приходит и может прийти в человеческий ум», — писал Декарт. А Лейбниц пытался заменить рассуждения вычислениями и превратить язык в своеобразную алгебру мысли.

Но все это были теоретические построения. Лишь с появлением электронных вычислительных машин от рассуждений философов появилась реальная возможность перейти к осуществлению их идей на практике.

### МЕЧТЫ, МЕЧТЫ...

7 января 1954 года в конторе фирмы «Интернейшел бизнес мэшин» в Нью-Йорке была проведена первая публичная демонстрация перевода с языка на язык — перевода, который впервые за всю историю человечества делал не сам человек, а его механический помощник. Электронная счетная машина ИБМ-701 перевела математический текст с русского языка на английский. На следующий год был осуществлен первый машинный перевод в нашей стране. С английского языка на русский переводила отечественная машина БЭСМ.

Затем в различных странах мира с помощью машин были сделаны переводы с французского, немецкого, китайского, японского языков. Встал вопрос о создании единого языка-посредника, на который и с которого можно было бы переводить любой язык мира. Ведь такой язык гораздо выгодней, чем составление отдельных программ перевода с каждого конкретного языка на другой конкретный язык (для двух языков нужны две различные программы перевода, скажем, с русского на английский и с английского на русский; для четырех языков таких программ нужно будет двенадцать, для двадцати — около трехсот, а так как число языков мира достигает нескольких тысяч, то число переводческих программ с каждого языка на каждый достигнет астрономически больших величин).

В качестве языка-посредника предлагались самые распространенные языки мира, например английский или русский. Были проекты использовать какой-либо из

искусственных международных языков, вроде эсперанто (ныне разработано несколько сотен подобных языков-посредников). Ленинградский лингвист Н. Д. Андреев предложил включать в язык-посредник только самые типичные, самые частые грамматические правила и слова, общие большинству языков мира. Такой язык-посредник будет представлять некое статистическое среднее этих языков. Есть и другая точка зрения: язык-посредник — это лишь система соответствий между различными языками мира, равнозначных друг другу слов и словосочетаний, а в материальной форме такой язык может и не существовать...

Проблема создания языка-посредника оказалась тесно связанной с не менее важной проблемой специального языка для информационно-логических машин. Потребность в таких машинах с каждым годом ощущается все острее. Ибо с каждым годом неудержимо нарастает лавина информации, которая обрушивается на современного человека. Особенно это относится к науке и технике. В конце концов мы можем и не знать о том, как сыграли футболисты «Спартака» или сколько градусов мороза нынче в Антарктиде,— ущерб от этого невелик. А вот ущерб от незнания научных или технических новинок может исчисляться миллионами рублей и годами потраченного напрасно труда. Читать же всю литературу, выходящую даже по его прямой специальности, инженер или ученый в наши дни просто не в состоянии.

«Человек так же неспособен выучить наизусть все книги, хранящиеся в библиотеке, как и взвалить их себе на спину»,— очень метко сказал академик В. М. Глушков. И не только выучить, но и просто-напросто прочитать. Число научно-технических журналов в наше время уже превысило пятьдесят тысяч и продолжает расти. Добавьте к ним еще семьдесят пять тысяч книг по науке и технике, издаваемых ежегодно. Плюс более трехсот тысяч описаний к авторским свидетельствам и патентам. Плюс сотни тысяч ежегодных научно-технических отчетов. И все это море информации кодируется не на одном, а на многих языках мира, причем и число таких языков науки и техники постоянно увеличивается. Если десяток лет назад было достаточно знания европейских языков, то теперь ценная научно-техническая информация публикуется на арабском, хинди, японском, турецком, персидском языках...

Обуздать Ниагару книг, потоки информации с помощью машин — машин-переводчиков или информационно-логических машин, хранящих в своей электронной памяти сведения из самых различных областей знания; создать общий язык для машины-переводчика и машины-информатора, некий универсальный код науки; научить электронный мозг не только хранить и выдавать по требованию информацию, но и обрабатывать ее, делать новые выводы, умозаключения, сопоставления; включить ЭВМ как мощнейший усилитель нашего мышления в цивилизацию XX столетия — таковы были грандиозные и благородные идеи, которыми руководствовались кибернетики, лингвисты, социологи, логики после первых опытов машинного перевода.

Казалось, время осуществления этих идей не так уж далеко. Пройдет каких-нибудь десять—двадцать лет, и начнется одна из самых крупных научно-технических революций в истории человечества...

Идея перевода с помощью ЭВМ была высказана в 1949 году. Со времени первого публичного перевода машиной с языка на язык прошло почти четверть века. Что же происходит в наши дни, в последней четверти двадцатого столетия?

### ...ГДЕ ВАША СЛАДОСТЬ?

Нет, кибернетическая революция, о которой мечтали четверть века назад и которую связывали прежде всего с машинным переводом и созданием информационно-логических машин, до сих пор не произошла. ЭВМ не переводят с языка на язык ни технические, ни тем более научные тексты. И хранителями знаний по-прежнему служат полки библиотек, а не блоки электронной памяти машин.

Более того. По свидетельству самих специалистов по машинному переводу, в наши дни свое основное внимание они сосредоточивают не на переводе, не на машине и не на алгоритме, формальной программе перевода!

В чем же тут дело? В порочности самой идеи, будто машина может переводить с языка на язык? Нет, идея эта подавляющим большинством ученых признается верной.

Тогда, быть может, все дело в технических трудностях: слишком маленькой оперативной памяти ЭВМ, недоста-

точной скорости, трудности ввода информации? Опять-таки все эти проблемы не имеют отношения к реальным проблемам машинного перевода.

На исследования в этой области отпускались большие средства, над машинным переводом работали целые лаборатории. Случалось, что попадали в них люди, далекие от науки, видевшие в МП лишь легкую поживу. Но ведь в основном и в нашей стране, и за рубежом проблемой автоматического перевода занимались крупные ученые, причем и математики, и лингвисты, и логики... И все-таки проблема эта не решена до сих пор.

Почему? Да потому что обманчивая простота механизации перевода вступила в совсем необманчивую сложность нашего человеческого языка. Чтобы сделать машинный перевод не проблемой, а реальностью, нужно сделать науку о языке достаточно точной, иначе нельзя перевести ее положения на формулировки программ ЭВМ. Вполне понятно, что ни десяти, ни двадцати лет не хватит для такой коренной перестройки. Да и вообще возможна ли она до конца — это тоже еще под вопросом.

Ну, а как же первые переводы с помощью машины? — спросит читатель. Ведь переводила же ЭВМ с русского на английский, с английского на русский и т. д.?

Переводила, это бесспорный факт. Но вот что она переводила — это другое дело. Есть такая шутка: «Черное, с крыльями, жужжит, жук — что это такое?» Ответ однозначен — жук, потому что он содержался в самом вопросе. Нечто подобное было и в первом опыте публичного машинного перевода, осуществленного в рекламных целях фирмой ИБМ. Текст был тщательным образом отпрепарирован, предельно упрощен. Правила грамматики также были элементарны. И словарный запас минимален.

Когда же стали увеличивать объем словаря, усложнять грамматику и пытаться перевести не препарированные, а подлинные тексты, перед исследователями стали задачи, неразрешимые на уровне современных знаний о языке. Проблем было сотни: слова, имеющие несколько значений, и слова-омонимы; грамматические правила и неизбежные в любом языке исключения из правил; многообразие не только слов, но и грамматических форм. В русском языке насчитывается сто пятьдесят семь различных глагольных окончаний, в немецком — триста

пятьдесят четыре, а в языке аранта — около тысячи (правда, пока что аранта не имеют письменности и поэтому проблема машинного перевода для этого языка Австралии неактуальна).

Перечень подобного рода проблем можно было бы продолжить, но вам, пожалуй, ясна наша главная мысль: язык оказался слишком сложен для машины, а лингвистика — слишком «гуманитарна», чтобы изложить накопленные ею сведения о языке в виде формул и алгоритмов. Больше того, даже для информационно-логических машин, которые, казалось бы, должны иметь дело со строгими и однозначными терминами науки, по сей день не удалось найти приемлемого кода, языка-посредника между информацией, накопленной человечеством, и электронной памятью ЭВМ. А причина этого все та же: наш человеческий язык. Ведь именно на этом языке «закодирован» весь океан современной информации — научной, технической и просто житейской.

Возьмем слово *белок*. Как закодировать его? Химик понимает под белком одно, биолог — другое, демограф — третье, повар — четвертое, врач-окулист — пятое, генетики — шестое и т. д. Какое из этих значений мы должны вводить в машину? Или ограничиться структурной формулой, генетическим кодом, то есть заменить слово условными знаками той или иной науки?

Но, во-первых, не для всех наук мы имеем такие знаки. Во-вторых, с помощью знаков можно описать далеко не все положения науки, которые мы выражаем средствами обычного языка. В-третьих, в большинстве областей науки и техники термины — это не просто условные словесные знаки, а все-таки слова, они связаны с другими словами языка ассоциациями, хотим мы этого или не хотим (лингвисты говорят даже об эмоциональности терминов!). Превращая слово-термин в кодовый знак, мы можем отсечь от него нечто существенное, важное.

Как же быть? Есть ли выход из сложившейся ситуации? Или непреодолимая пропасть разделяет наш язык и алгоритмический язык машин?

## ИНФОРМАТИКА — НАУКА ДЕЛОВАЯ

Информационным взрывом называют иногда процесс, начавшийся в середине нашего столетия. Научно-технический прогресс — причина этого взрыва, лавино-

образно нарастающего потока информации в самых различных областях науки и техники. Укротить эту лавину необходимо. Только в нашей стране трудится около миллиона научных работников; число их на земном шаре гораздо больше. И все эти миллионы людей, несмотря на их старания и желания, не могут работать с полным КПД своего мозга. Ибо не в состоянии прочитать литературу по своей специальности, выходящую в мире.

«Если бы химик, свободно владеющий 30 языками (условие невероятное), начал с 1 января 1964 г. читать все выходящие в этом году публикации, представляющие для него профессиональный интерес, и читал бы их по 40 часов в неделю со скоростью 4 публикации в час, то к 31 декабря 1964 г. он прочитал бы лишь  $\frac{1}{20}$  часть этих публикаций», — говорил академик А. Н. Несмеянов. А ведь с 1964 года число публикаций по химии из года в год увеличивалось, и сейчас наш химик-полиглот не прочел бы и одной тридцатой всей литературы по специальности.

Поток публикаций возрастает, а вместе с тем ежегодно возрастает и необходимость чтения этих публикаций. Все чаще и чаще обращаются к научным публикациям инженеры и техники. Их в мире не миллионы, а десятки миллионов. Больше века прошло со времени открытия электрического тока Гальвани до создания первой электростанции. Полстолетия потребовалось телефону, чтобы из научного открытия воплотиться в техническое изобретение. Но такие темпы характеризовали XVIII и XIX столетия. В нашем веке с момента открытия деления ядер урана до создания атомного реактора прошло лишь три года, а до запуска первой атомной электростанции — пятнадцать лет. Научные открытия используются сейчас почти моментально, если мерять время темпами прошлых лет. Зато, в отличие от прошлого, поиск нужной информации в океане книг, статей, журналов, патентов отнимает уйму времени.

И не только времени, но и денег. Ежегодно на документальный поиск в США тратится свыше миллиарда долларов. И тем не менее в тех же США — на неоправданное дублирование тратится не менее десяти процентов всех средств, что отпущены на научно-исследовательские и конструкторские работы. Вот два характерных примера. На эксперимент по засеву облаков было израсхо-

довано более четверти миллионов долларов. Результаты этого эксперимента появились в публикации, но затерялись в потоке отчетов и статей. И вскоре эксперимент был повторен, причем дублирование его обошлось уже в три миллиона долларов. Национальное управление по авионавигации и исследованию космического пространства потратило год и восемнадцать миллионов долларов на разработку ракетной системы «Атлас-Вега». Одновременно столько же времени и средств ушло на разработку аналогичной системы «Атлас-Агента В», которой занималось министерство ВВС Соединенных Штатов.

Вполне понятно, что на службу информации различные фирмы, исследовательские институты, министерства и управления зарубежных стран не жалеют средств — эти средства окупаются сторицей.

В нашей стране информационная служба централизована. О создании единой общегосударственной системы сбора и обработки информации для учета, планирования и управления записано в «Основных направлениях развития народного хозяйства СССР на 1976—1980 годы», утвержденных XXV съездом КПСС. Совет Министров СССР принял специальное постановление, возложив на Государственный комитет по науке и технике руководство научно-технической информацией в нашей стране, определение дальнейшего развития системы этой информации, координацию исследовательских и технических работ, методическое руководство всей сетью научно-технической информации, разграничение функций между отдельными органами информации и контроль над их деятельностью.

Сеть информационных органов СССР включает такие специальные институты, как ВИНТИ — Всесоюзный институт научной и технической информации, где работают тысячи штатных сотрудников и десятки тысяч переводчиков, ЦНИИПИ — Центральный научно-исследовательский институт патентной информации, ВНИИКИ — Всесоюзный научно-исследовательский институт научно-технической информации, классификации и кодирования. В каждой отдельной отрасли, как правило, существуют свои институты, а отделы или бюро научно-технической информации есть почти во всех солидных научно-исследовательских и проектно-конструкторских институтах, на предприятиях промышленности, сельского хозяйства, транспорта и связи.

Отпустить необходимые средства, создать единую сеть службы информации — это, так сказать, задача общественная, социальная, государственная. И она в нашей стране решается. Но есть еще задачи сугубо внутренние, связанные с самой спецификой проблемы. Решить их должна особая наука об информации, или, как ее называют ныне, информатика (известный советский логик В. А. Успенский предложил термин «документалистика», однако в нашей литературе первый термин получил гораздо большее распространение).

«Информатика — это научная дисциплина, изучающая структуру и свойства (а не конкретное содержание) научной информации, а также закономерности научно-информационной деятельности, ее теорию, историю, методику и организацию,—пишут А. И. Михайлов, А. И. Черный и Р. С. Гиляровский в монографии «Основы информатики». — Целью информатики является разработка оптимальных способов и средств представления (записи), сбора, аналитико-синтетической переработки, хранения, поиска и распространения научной информации. Информатика имеет дело со смысловой (семантической) информацией, но не занимается качественной оценкой этой информации. Такая оценка может производиться лишь специалистами в соответствующих областях науки или практической деятельности».

Информатика находится в начале своего становления. Как и всякая другая наука, она сначала собирает факты и обобщает их, затем анализирует и выявляет общие закономерности. И, как считают все специалисты, основная теоретическая задача информатики в наши дни — это выяснение закономерностей, по которым происходит создание семантической, смысловой информации, передача этой информации и практическое использование в самых различных сферах человеческой деятельности, начиная с доказательства математических теорем и кончая применением достижений науки и техники в повседневном быту.

Содержание, семантика, смысл кодируется с помощью специальных языков науки и техники и прежде всего с помощью нашего человеческого языка, сложного и неоднозначного. Перевод с этого языка на язык информатики — так можно сформулировать задачу номер один, которая стоит перед современными исследователями.

Язык машины — язык чисел. На этот язык в принципе должна быть переведена вся информация, накопленная человечеством в области науки и техники, вообще вся сумма знаний, запечатленных в печатной продукции. Кодирование числами началось задолго до появления ЭВМ и информационного взрыва. Списки книг и произведений составлялись и древними греками, и египтянами, и жителями Двуречья. Уже в VII веке до н. э. на глиняных табличках из библиотеки ассирийского царя Ашшурбанипала давалось заглавие, номер таблички или собрания табличек, образующих «глиняную книгу» и т. д. Более ста лет назад, в 1876 году, американский библиотеквед Мелвил Дьюи предложил классифицировать книги, относящиеся к различным областям знания, с помощью десятичных чисел.

Каждый основной раздел этой классификации обозначался одной цифрой, от нуля до девяти. При дальнейшем его делении к этой цифре присоединяется вторая, а затем и третья цифра. А чтобы индексы были всегда трехзначные, к однозначным и двузначным числам Дьюи добавлял нули. Например, естественные науки получают индекс 500, математика тогда будет под индексом 510, астрономия — 520, физика — 530, химия — 540, геология — 550, палеонтология — 560 и т. д.

Сведения о той или иной сфере могут быть в периодических изданиях, словарях и других публикациях. Они могут касаться истории вопроса, могут говорить о практическом применении и о многом другом. Дьюи разработал список таких делений, занумеровал его, и теперь тот или иной номер может присоединяться к трехзначному числу, индексу любого раздела классификации. Например, физика имеет индекс 530, а ее раздел механика — индекс 531. Тогда словари по механике кодируются как 53103 (словари обозначены в списке Дьюи числом 03), история механики — 53109 (09 — обозначение «истории вопроса»).

А как быть со странами или языками? Дьюи предложил для обозначения их использовать окончания индексов филологии и истории, которые зафиксировали различные языки и страны мира. И тогда геология Европы получает обозначение 55040 (первые три цифры —

индекс геологии, две вторые—Европы), геология Азии—55050, геология Африки — 55060 и т. п.

Десятичной классификацией Дьюи пользуются почти девяносто процентов библиотек США и Великобритании. Однако в большинстве стран мира, в том числе и в нашей стране, используется другая десятичная классификация, именуемая универсальной, сокращенно УДК. Создана она была в начале нашего столетия и с тех пор продолжает совершенствоваться и расширяться. В последних изданиях УДК содержится более ста тысяч рубрик, охватывающих самые различные области человеческого знания. Полный объем ее таблиц составляет около пятисот авторских листов, то есть десяток томов по триста — пятьсот страниц в каждом.

Тысячи учреждений почти в сотне стран мира пользуются системой УДК. По сути дела, это своеобразный международный язык-посредник. Ведь тексты, выходящие на множестве различных языков мира, индексируются с помощью числового кода УДК. Зная этот код, мы можем перевести его символы средствами своего родного языка.

Система УДК представляет собой иерархию. Вершина ее — десять цифр, от нуля до девяти, обозначающих главные разделы: 0 — общий отдел, 1 — философия, 2 — религия, 3 — общественные науки и т. д. Далее по тому же десятичному принципу каждый из разделов дробится на отдельные подразделы, те, в свою очередь, на группы, группы — на подгруппы и так до тех пор, пока сохраняется необходимость членения. Чем больше развита та или иная область знания, тем больше и глубина деления.

Например, в минералогии и кристаллографии она достигла десятой степени, то есть отдельные понятия записываются в виде десяти цифр (структура кальцита обозначается числом 548.736.442.2, где цифра 5 обозначает математику и естественные науки, 4 — химию и т. д.). Там, где надо, глубина может увеличиваться безгранично.

Приведем для образца запись на УДК фрагмента текста на русском языке. Фраза «Строительство из стали с экономической точки зрения» будет записана так: 624.94.003.1. Первая цифра, шестерка, кодирует прикладные знания, медицину, технику; вторая — двойка — технику и инженерное дело. Сочетание цифр 624.94. обо-

значает строительство, каркасные конструкции. А цифры 003.1 — определитель «экономической точки зрения».

И все-таки какой бы совершенной ни была УДК или любая другая классификация этого типа, она не решает главных проблем, возникших в связи с информационным взрывом. Ибо такие системы, если можно так выразиться, одномерны. А поиск информации в наши дни идет по самому различному набору признаков, в многочисленных «измерениях». Специалисты по информатике иллюстрируют это на убедительном и наглядном примере.

Допустим, мы размещаем книги библиотеки по цвету их переплета. В одну группу помещаем книги в красном, в другую — в желтом, в третью — в голубом переплете. Но вот к нам попадает книга в зеленом переплете. Куда ее поместить? Мы выделяем еще одну группу. Когда у нас будет слишком много книг в светло-голубых переплетах, мы эти книги выделим в отдельную подгруппу или даже группу. Оттенки спектра бесчисленны. И мы можем, если потребуется, дробить и дробить эту группировку, используя тот или иной цвет или его оттенок.

Именно так и поступает УДК или подобная ей система. Только вместо бесконечного спектра здесь используется бесконечный ряд чисел. Но представим, что в нашу библиотеку попадает книга, переплет которой частью красный, а частью голубой. Куда ее отнести? К голубым или красным? Если мы отнесем ее и к голубым, и к красным книгам, получится двусмысленность, да и неточность: ведь книга-то не красная и не голубая, а красно-голубая.

Между тем в наше время то и дело возникают именно такие вот «красно-голубые», смешанные области знания, возникшие на стыке наук. Практические же применения их, если продолжить наше сравнение, вообще «серо-буро-малиновые»: медицина переплетается с электроникой, психологией, математикой. Лингвистика, как вы сами убедились, читая эту книгу, стыкуется с инженерией, статистикой и т. д. и т. п. УДК и другие подобные ей системы малоприспособны для информационного поиска по любым, заранее не предусмотренным сочетаниям предметов. И совсем непригодны они для поиска по единичным предметам, а также предметам межотраслевого характера, рожденным стыком наук или

неожиданным контактом техники и нового открытия в науке...

Это вывод современных специалистов по информатике. Но задолго до рождения этой науки, еще в 1933 году, выдающийся индийский ученый Шиали Рамамрита Ранганатан указал на слабости цифровой классификации по типу УДК, обозначения «предметов и книг порядковыми числами, необходимого для достижения специфических целей». И не только указал на недостатки старой системы, но и разработал свою систему, весьма оригинальную. Принципы ее и по сей день используют ученые, работающие в области ИПЯ — информационно-поисковых языков. В наши дни создано несколько тысяч таких ИПЯ, так что по своему количеству они могут соперничать с естественными языками.

### **«ПУСТО — НЕПУСТО», БИТ, «КОРАН»...**

Рассказ о различных информационных языках потребовал бы не очерка, а целой книги. Мы ограничимся лишь несколькими ИПЯ, наиболее известными или интересными.

«Пусто — Непусто» — так назывался первый в нашей стране информационный язык, разработанный в ВИНТИ в начале шестидесятых годов. Этот ИЯ предназначался для поиска рефератов по электротехнике с помощью ЭВМ. Знаменательные слова, или, говоря языком информатики, дескрипторы кодировались трехзначными числами. Например, фраза «данные о напряжении тока» на этом языке записывалась так: 153.414.883 (153 — данные, 414 — напряжение, 883 — ток). Информационный язык «Пусто — Непусто» применен был для текстов на двух языках — русском и английском.

В 1965 году В. И. Тарасовым был предложен информационный язык «Кристалл». В отличие от «Пусто — Непусто» кодирование в нем смешанное, с помощью слов и цифр. Слова текста записываются в исходной форме, без грамматических форм. С помощью чисел обозначаются отношения между словами. В наши дни язык «Кристалл» применяется в самых различных сферах: с его помощью отыскивается информация в области химии, медицины, приборостроения, легкой промышленности и т. д. Свыше ста тысяч различных документов переведено на этот информационно-поисковый язык.

Для того чтобы облегчить поиск публикаций по органической химии с помощью машин, разработан информационный язык, получивший название «ИЯ стандартных фраз». Специфика этого языка в том, что он имеет дело не с простыми текстами, представляющими линейную последовательность букв, а с химическими формулами. Формулы же органической химии, как известно, структурны, они располагаются в двумерном пространстве. Поэтому пришлось разрабатывать особый код, с помощью которого двумерные формулы превращались в линейную последовательность цифр.

Наиболее сложный и четкий по своей логике ИЯ был разработан в нашей стране группой сотрудников Института кибернетики АН УССР под руководством Э. Ф. Скороходько. Язык этот получил наименование БИТ. Термины на нем могут быть представлены в трех вариантах: в виде формул, графиков и таблиц. Первый вариант записи наиболее экономичен, второй — нагляден, третий — пригоден для машинной обработки текста (киевские кибернетики разработали специальные программы, которые позволяют ЭВМ переводить текст с естественного языка на искусственный язык БИТ).

В основе этого информационного языка лежит мысль о том, что структура внешнего мира состоит из единиц двух видов: предметов и их отношений. Эти первичные единицы порождают производные — ситуации. Предметы обозначаются символом  $X$ , который может иметь различные индексы вверху и внизу. Так, термин «изображение» получит символ  $X_{024}^0$ , где верхний нолик означает, что мы имеем дело с элементарным понятием, а нижнее число показывает номер данного термина в списке. Более узкий термин, например «полупроводниковый триод», закодируется в виде символа  $X_{301}^2$ .

Отношения между предметами — релатемы — кодируются латинской буквой  $R$ , которой также приписаны верхние и нижние индексы. Ситуации на языке БИТ записываются в виде предложений, состоящих из соединения символов  $X$  и  $R$ , имеющих различные индексы. Словарь терминов в этом языке разработан для текстов по радиоэлектронике, но, судя по всему, БИТ может дополняться новыми терминами и понятиями.

Естественно, что появление нового предмета требует нового индекса. Для отношений это не обязательно:

производные релатемы могут образовываться сочетанием основных (например, сочетание релатемы «быть местом» и «иметь субъект» образует релатему «быть местом действия»).

Информационный язык БИТ — это усовершенствованный вариант так называемого Кливлендского семантического кода. Код разрабатывался группой американских ученых в городе Кливленд для механизированного поиска информации по металлургии. Алфавитом кода служат не только латинские буквы, но и специальные знаки и числа от 000 до 999. Каждое знаменательное слово, то есть дескриптор, записывается в виде сочетания трех согласных букв. После второй согласной оставляется пробел, который заполняется какой-либо гласной. Вставная буква указывает на определенное отношение, например: А — это «принадлежность, быть предметом»; О — «производить» и т. п. Из простых дескрипторов строятся составные, более сложные. Так, для записи слова *телефон* используются дескрипторы *прибор, передача, электричество, информация* (ведь телефон — это и есть прибор, предназначенный для передачи информации при помощи электричества).

Таким образом, мы имеем дело с набором элементарных понятий, или, как называли их создатели Кливлендского семантического кода, смысловыми множителями — по аналогии с физическими базисными величинами (хотя, конечно, никакого умножения тут не происходит, лишь из кирпичиков простых дескрипторов строятся более сложные).

Многие специалисты по информатике считают, что в первую очередь надо создавать не всеобщие ИЯ, а специализированные, нацеленные на какую-то одну область знания. Таков, например, язык «Коран», предназначенный для поиска сведений, содержащихся в священной книге мусульман. Словарь ИЯ «Коран» содержит около пятисот терминов, набитых на перфокарты с указанием, в каком месте священной книги встретилось то или иное слово. Любой отрывок из писания мусульман может быть закодирован на языке «Коран» в виде условных символов и ключевых слов: так,  $d_1$  обозначает Аллаха,  $d_4$  — ангелов,  $d_5$  — демонов,  $d_6$  — джиннов и т. д.

Специальные ИЯ создаются также для записи неязыковой информации. Например, для поиска картографических материалов, различных объектов, интересных для

археологов (ваз, монет, мозаики, орнамента, оружия), для классификации товарных знаков по изображению, цвету, геометрической форме.

## ТЕЗАУРУС — ЭТО СОКРОВИЩЕ!

Мы начали наш рассказ с Раймонда Луллия, жившего в эпоху, весьма далекую от эры ЭВМ, однако предвосхитившего идею механического «усилителя разума». В том же XIII веке была выпущена книга, давшая название теме, которая является самой актуальной в создании информационных языков (а те, в свою очередь, — ключ к решению основной задачи информатики). Речь идет о создании так называемых тезаурусов.

Термин этот происходит от греческого слова, имеющего значение: сокровище, богатство, клад, запас. От греков это слово попало и в латынь, где стало обозначать сокровищницу, кладовую, склад. Флорентиец Брунетто Латини, живший в XIII веке, назвал «Книгой о сокровище» свою систематизированную энциклопедию знаний той эпохи. В XVI столетии отец и сын Этьенны выпустили толковый словарь языка древних греков, озаглавленный «Сокровищница греческого языка». Этот труд не утратил своего значения и поныне.

А вслед за тем выходят тезаурусы-словари, которые являются своего рода «перевернутыми» толковыми словарями. В толковых словарях вроде знаменитого словаря Даля слова расположены в алфавитном порядке. Значения их раскрываются примерами, взятыми из обыденной речи, из художественной литературы, из народных говоров. В тезаурусах слова и выражения систематизированы не по написанию или произношению, а по их значению. Недаром же тезаурусы называют еще идеологическими словарями.

В 1808 году был выпущен тезаурус санскрита. Все слова в нем объединялись в классы: Небеса, Погода, Время, Добродетели, Пороки, Ад и т. д. А те, в свою очередь, делились на подклассы (например, Ад подразделялся на Ад как таковой, Души умерших, Страдание, Боль).

«Нельзя ли подобным образом рассортировать по смыслам слова не мертвого санскрита или древнегреческого, а живого английского языка?» — такой мыслью задался П. М. Роджет. Ведь подобный словарь принес бы

большую пользу в смысле «облегчения выражения понятий», а также «при написании сочинений».

Роджет выпустил свой «Тезаурус английских слов и фраз» в 1852 году. С той поры его словарь выдержал около сотни изданий. Тезаурус Роджета делит значения всех слов, входящих в него, на шесть основных категорий: абстрактные отношения, пространство, материя, разум, воля, чувственные и моральные силы. Категории эти подразделяются на двадцать четыре подкласса, те — на восемьдесят шесть подподклассов и т. д. Всего тезаурус включает тысячу различных категорий, по которым и разносятся значения слов. Если слово имеет несколько значений, то оно попадает в несколько рубрик. Например, слово knot входит в группу 45 (связь), а также 59 (запутанность), 72 (группа), 219 (пересечение), 249 (округлять), 321 (плотность), 704 (трудность) и т. д.

Словарь Роджета при многочисленных своих переизданиях не раз и перерабатывался. Но общая схема его не менялась, добавлялись лишь новые слова и понятия. Между тем этой схемой невозможно охватить необыкновенно расширившуюся в нашем веке сферу науки и техники (напомним, что Роджет жил более ста лет назад). Вот почему, составляя современные тезаурусы, ученые стремятся избежать ограниченности словаря Роджета.

Наиболее полным словарем такого типа считается идеологический словарь немецкого языка, составленный Францем Дорнзайфом. В нем уже не шесть, как у Роджета, а двадцать основных категорий, каждая из которых делится на несколько десятков подклассов. Подклассов может быть от десяти до девяноста, но зато они являются последней рубрикой. В словаре Дорнзайфа нет подподклассов, излишне усложнявших тезаурус Роджета. (Так, первый класс Роджета *абстрактные отношения* делился на 1.1. Бытие. 1.1.1. Абстрактное. 1.1.1.1. Бытие. 1.1.1.2. Небытие. 1.1.2. Конкретное. 1.1.2.1. Вещественность. 1.1.2.2. Невещественность и т. д. Нетрудно запутаться во всех этих рядах единиц, точек и двоек.)

Сразу после заглавного слова, опять-таки в отличие от тезауруса Роджета, даются ссылки на другие классы и подклассы, что очень удобно для работы со словарем. Вот как выглядит самая краткая из словарных статей

класса под номером тринадцать (Знак, Сообщение, Язык): 13.17. Значение. (см. 9.44. Значимость. 12.31. Понимать. 13.44. Объяснение. 13.53. Перевод. 14.12. Указатель слов). Это значит. Значит, Хочу сказать. Именно. Итак. В этом смысле. Собственно. Буквально. Выражать. Значить. Свидетельствовать. Означать. Обозначить. Иметь целью. Намекать. Извещать. Показывать. Иметь в виду. Относиться к. Служить для. Покрывать. Семантически. Семасиологически. Выражение. Значение. Содержание. Дух. Понимание. Значение слова. Смысл слова. «Синоним». Наука о значении. Семантика. Семасиология.

В словаре Дорнзайфа приведена классификация царства животных и растений, что создана современной наукой, дана систематика минералов, разработанная геологией, и систематика болезней, принятая медициной. При переиздании этот тезаурус пополняется новой лексикой (так, в последнее издание введены слова *спутник*, *водородная бомба* и т. д.). И все-таки как бы ни был полон словарь Дорнзайфа, он не в состоянии охватить всей поистине необозримой терминологии, что существует в современной науке и технике. Да и отношения между смыслами слов слишком общи для того, чтобы можно было вести направленный поиск информации по ним. Вот почему с начала шестидесятых годов в различных странах мира создаются специальные тезаурусы, цель которых — охватить терминологию той или иной области науки и техники. И не просто охватить, но и служить путеводителем в поисках научно-технической информации.

В наши дни созданы десятки тезаурусов в самых различных сферах. Есть тезаурус Евроатома, тезаурус терминов по целлюлозе и бумаге, химико-технологический тезаурус, тезаурус по металлургии и т. д. Принципы их построения совпадают с принципами классических тезаурусов-словарей. Они, как говорят специалисты по информатике, предназначены как для нахождения разных способов выражения одной и той же мысли, так и для оптимизации решения информационно-поисковых задач. Только, в отличие от словарей Роджета и Дорнзайфа, специализированные тезаурусы стремятся не к всеобщему охвату значений, а к детальному и углубленному охвату терминологии той или иной области науки и техники.

## КАК НАПОЛНЯЮТ СОКРОВИЩНИЦУ?

Как составляются тезаурусы?

Роджет ничего не знал и, разумеется, не мог знать, каких результатов достигнут науки, занимающиеся проблемой смысла, но он интуитивно понял основные принципы составления тезаурусов: сначала собираются слова из текстов или различных словарей. Затем слова эти выстраиваются по алфавиту, а после из них образуются определенные группы, связанные по смыслу.

Какой может быть эта смысловая связь? Во-первых, одно и то же значение можно выразить с помощью разных слов или оборотов. Вполне понятно, что все синонимы будут объединены. Во-вторых, смыслы слов могут быть полностью противоположны (как антонимы типа *холодный — горячий, мертвый — живой* и т. п.) В-третьих, смысл одного слова может включать в себя смысл другого, или даже многих слов, например значение слова *человек* включает в себя наименования сотен различных национальностей нашей планеты. Смысловые связи между словами могут строиться по типу *причина — следствие* или *часть — целое*. Наконец значения слов между собой связаны самыми различными ассоциациями. Причем именно проблема ассоциаций является наиболее трудно разрешимой задачей при построении того или иного тезауруса.

Специалистам в данной области науки нетрудно выделить термины, которыми эта наука пользуется. Сложнее построить иерархию из слов-терминов по смыслу. Однако и эта задача разрешима логическим путем. Можно разграничить смысл слов и словосочетаний, которые образуют новые значения. Например, термины *скорость бурения* и *бурение скоростное* — это не синонимы, они имеют совершенно разные значения, так же как *дюнный песок* и *песчаная дюна*. Но специалисты по информатике с успехом ликвидируют подобного рода случаи (в первом случае бурение является аспектом, а во втором — атрибутом; в сочетании *песчаная дюна* песок — вещественная составляющая, а в сочетании *дюнный песок* — причина образования дюны).

С ассоциациями дело обстоит много сложнее. Ибо здесь действуют законы не двузначной, а многозначной логики.

В тезаурусах группируются, например, понятия *ра-*

нение — несчастные случаи — диагноз — медицина — патология — терапия или чернила — краски — печатание — письмо. Интуитивно нам понятно, что объединение этих слов оправдано. Но попробуйте-ка доказать его путем формальной логики, и вы убедитесь, что это далеко не так просто, как на первый взгляд кажется. А ведь такие связи существуют не только между обычными словами языка, но и научными терминами.

Еще в 1942 году американский исследователь А. Л. Болдуин попытался внести количественные меры в изучение смысловых связей между словами. Он подсчитал совместную встречаемость слов в письмах, написанных одной женщиной, и показал, что частоты их могут быть указателями настроений этой женщины, круга ее интересов и т. п. С тех пор предложено несколько десятков формул, по которым можно вычислять смысловые связи слов.

Применяются для выявления этих связей и электронные вычислительные машины.

Конечно, как замечает один из специалистов по информатике — Р. М. Нидхем, «нельзя ожидать, что в результате нескольких часов работы ЭВМ будет получено нечто, что могло бы сразу же заменить интуицию и труд ученого (который все равно необходим при выборе единиц анализа и интерпретации результатов). На что мы можем надеяться — так это на то, что выявятся группы слов, ранее не замеченные, но потенциально возможные».

Тезаурусы составляются учеными в «союзе» с машиной.

Сначала группа специалистов подчеркивает термины в данной области знания. Затем тексты вводятся в ЭВМ, которая выдает эти слова в исходной форме, то есть в именительном падеже, единственном числе и т. п. Потом специалисты разбивают на участки смысловое «поле», внутри этих участков производят более мелкое деление, а полученную схему оформляют в виде анкеты.

Анкета заполняется специалистами по каждому термину, а затем она поступает в ЭВМ, которая объединяет все термины, имеющие одинаковые признаки. И, наконец, человек составляет из них тезаурус.

Нельзя ли использовать тезаурусы в качестве языка-посредника при машинном переводе? Этой мыслью задалась английские ученые Кембриджской лингвистической группы.

В 1959 году один из крупнейших авторитетов в области машинного перевода, Бар-Хиллел, заявил, что высококачественный перевод с помощью ЭВМ принципиально невозможен. Ибо в нашем языке есть такая многозначность слов, которую нельзя ликвидировать, не обращаясь к внешнему миру, к внеязыковой, или, как говорят ученые, экстралингвистической ситуации, машине недоступной. Бар-Хиллел привел образец подобного рода принципиально непере译имой фразы... И в том же году сотрудники Кембриджской группы добились правильного машинного перевода этой фразы!

Если все предыдущие работы по машинному переводу стремились игнорировать значение, полагая, что ЭВМ недоступно «понимание» смысла текста, то в работах Кембриджской группы акцент делался именно на перевод значения. Ведь его-то передача и является целью всякого перевода с языка на язык. Сделать же этот перевод по смыслу помог знаменитый тезаурус Роджета.

Перевод с помощью тезауруса происходит не так, как обычный человеческий или стандартный машинный перевод. Это перевод не отдельных слов или фраз. Переводятся целые абзацы. Вначале строится единая «формула смысла» данного абзаца, причем смысл фраз, входящих в него, кодируется номерами тезауруса. Затем для полученной «формулы» смысла находят соответствующие «формулы смысла» того языка, на который переводится текст.

Работы подобного рода ведутся не только в Англии, но и в Италии, в Миланском центре кибернетики и лингвистики, а также в нашей стране. Несколько различаясь в деталях, методика везде одина: выявление своего рода «атомов смысла», семантических множителей, из которых слагается значение отдельных «молекул», слов языка.

Подобные семантические множители применялись, как вы помните, в информационном языке, созданном учеными из Кливленда (Кливлендский семантический код). Но там задача была проще: из набора элементар-

ных понятий формировались более сложные термины. Речь шла именно о терминологии, а не о языке со всеми его оттенками смыслов и значений слов. Здесь же задача, так сказать, на порядок сложнее.

Кембриджская группа пользовалась тезаурусом Роджета. К глубокому сожалению, такого словаря для русского языка до сих пор никто не составил. Правда, сейчас в нашей стране создается не менее интересный словарь, значение которого для машинного перевода трудно переоценить.

Словарь этот — толково-комбинаторный. В нем приводятся все сочетания русских слов, смыслы которых образуют новый, невыводимый из их сложения смысл. Например, проливной дождь, грибной дождь, косой дождь, слепой дождь и т. д. Если в обычных словарях разъясняется смысл отдельного слова, то в новом словаре толкуются словосочетания.

Работа по созданию такого словаря требует очень много труда. Ведь число различных русских слов равно сотням тысяч. А сколько же различных словосочетаний, имеющих «третий смысл», может быть в нашем языке! Вероятно, их не меньше, а гораздо больше отдельных слов и достигает миллиона. Составление словаря продлится не один год, а то и не один десяток лет (в настоящее время удалось зафиксировать сочетания двухсот слов, так что вся работа еще впереди). Поэтому сейчас исследователи стремятся выявить наиболее общие смысловые множители русского языка. С их помощью они надеются определить своеобразную «геометрию смысла», законам которой подчиняется и наш язык, и, по всей видимости, любые другие языки, так как любое содержание, высказанное на одном языке, может быть передано средствами любого языка (опыт переводчиков убедительно это доказывает: стихи Пушкина понимает и «друг степей калмык», и финн, и эвенк, и граждане США, ФРГ, Мексики и т. д.).

Приведем образец одного из параметров смысла, который удалось выявить в русском языке. Условно он кодируется как Magn, от латинского magnus, то есть высокий, главный, большой, обширный и т. п. Параметр этот определяет слова, имеющие значения высокой степени. Точнее, наш Magn определяет тот или иной оттенок смысла слова, к которому он относится. Возьмем, например, слово *друг*. Друг может быть близкий,

верный, большой, до гробовой доски, сердечный, закадычный, испытанный, старый... Синонимы? Разумеется, нет. Мой старый друг отличается от сердечного друга и от друга до гробовой доски. Но чем, собственно говоря, отличается?

Вот тут-то и начинается наша «геометрия смысла», определение в языке тех координат, которые позволяют нам в нашей жизненной практике считать друга Виталия закадычным, друга Евгения сердечным, а друга Володю испытанным. Как же выявить эти координаты? Очевидно, необходимо обратиться к носителям смысла, носителям языка, то есть к живым людям. И к текстам, которые пишутся на русском языке и понятны этим людям.

Вот какие оттенки удалось выявить в русском языке исследователям из лаборатории машинного перевода при Московском государственном педагогическом институте иностранных языков имени Мориса Тореза. Прежде всего «Магн вообще» делится на три подкласса. Первый — нулевой — подчеркивает значение слова, к которому он относится, вне зависимости от субъекта или объекта. Ограбление может быть дерзким или крупным, любить можно горячо или страстно. Второй подкласс подчеркивает усиление, относящееся к субъекту. Выражение «бежать быстро» нейтрально, оно относится к нулевому подклассу. А вот выражение «бежать изо всех сил» указывает на то, что бегущий изо всех сил старается использовать все свои возможности при беге (и тем не менее, как бы мы с вами ни старались бежать изо всех сил, бегущий Валерий Борзов нас легко опередит, хотя он будет бежать не изо всех сил, а просто быстро!).

Третий подкласс, в отличие от предыдущего, относится не к субъекту, а к объекту. Уничтожение может быть массовым, или тотальным; любовь — всеобъемлющей, капитуляция — полной, авторитет — общепризнанным, горе — народным, а враг — жестоким.

Однако в зависимости от степени или меры внутри каждого подкласса можно выделить несколько категорий или элементарных смыслов. Это — очень, большое количество, активность, длительность, выделение в своем классе. Чтобы каждый человек, владеющий русским языком, смог понять то, что формулируется на языке терминов (но термины тоже опираются на языковую интуицию носителей языка, именно она помогает нам

понимать друг друга безо всяких лингвистических теорий), приведем несколько примеров.

«Очень» указывает на высокую степень того или иного признака предмета, но без сравнения его с другими предметами, которые обладают теми же признаками. Здоровье может быть крепким, и дружить можно крепко, а интересоваться — глубоко. Вторая категория — «большое количество» — указывает на увеличение объема. Выбор может быть широким, информация — обширной, ограбление — групповым, вред — огромным. Третья — «активность» — связана с интенсивностью того или иного процесса. Работать с душой можно и не активно, а вот работать не покладая рук — это значит трудиться активно. Выражение «огромная энергия» подчеркивает количество этой энергии, а «кипучая энергия» — ее активность.

Категорию «длительность» легко понять, если в качестве примера привести выражения «продолжительные аплодисменты», «старая дружба», «смеяться до упаду», «любовь до гроба», «древние традиции» и т. п. Наконец, последняя рубрика — «выделение в своем классе» — указывает на то, что смысл выделяется из такой же смысловой группы особым подчеркиванием: оборудование бывает уникальное, урожай — рекордным, красота — необыкновенной и бежать можно изо всех сил...

### яп, мп, ия

Вероятно, теперь вы и сами поняли, как трудно, а порой просто невозможно описать в терминах, понятных ЭВМ, всю нашу сложную и нелинейную систему значений. И тем не менее система существует, иначе человек не смог бы понимать человека и ни один ребенок не смог бы обучиться родному языку, не говоря уже об иностранных.

Что же получается в итоге? Человек, решив передать машинам механическую работу, обнаружил, что, казалось бы, самые обыденные дела, вроде разговора, требуют таких сложных алгоритмов, таких тонкостей, о которых не догадывались даже лучшие умы прошлых веков.

И все-таки машины нужны для перевода! Современный мир, говорят ученые, сам по себе напоминает огромную переводческую машину, работающую со все

увеличивающейся скоростью. Каждый день с печатных станков сходит около шестидесяти переведенных книг, открывается три или четыре международных встречи, выпускается на экран несколько дублированных фильмов; пишущие машинки печатают, а дикторы оглашают несчетное количество переведенных страниц; бесконечное число документов переводится в самом разнообразном виде с одних языков на другие.

Но для того, чтобы включить в гигантскую «переводческую машину», которой является наш современный мир с его несколькими тысячами языков, ЭВМ, мы должны прежде всего разобраться, хотя бы приблизительно, в работе нашей собственной «вычислительной машины» — мозга, причем под определенным углом зрения — лингвистическим. Не инженер, а лингвист должен найти формулы языка, алгоритмы нашего повседневного чуда — языка, которое предстает перед нами каждый день и каждый миг. Задача же эта необычайно сложна и трудна.

Причем на пути ученых встают трудности самого различного характера: технические, лингвистические, логические. Предположим, что нам удалось решить проблемы смысловых множителей, разложить любое слово на отдельные «атомы смысла», его составляющие. Сколько же слов понадобится вводить в память машины?

Мы уже говорили, что словари таких развитых языков, как русский или английский, включают около полумиллиона слов. Но ведь есть еще и так называемые фразеологизмы, неперебиваемые буквально выражения, идиомы. Их в языке много тысяч. Добавьте к ним также десятки тысяч специальных терминов, в словари литературного языка не вошедших. А такой развитой терминологией обладают все науки, будь то химия, медицина, электроника или сама лингвистика («Словарь лингвистических терминов» О. С. Ахмановой включает семь тысяч единиц, а в него вошли далеко не все термины современной науки о языке, например, нет в словаре ни «инженерной лингвистики», ни «нейролингвистики», то есть целых дисциплин, имеющих свою собственную терминологию!). Так что общее число слов будет, пожалуй, превышать миллион. А ведь к ним надо прибавить еще правила грамматики и программы, объем которых составит не менее трети машинной записи словаря!

Это, так сказать, трудность чисто техническая. Не за горами то время, когда ЭВМ будут обладать памятью, достаточно большой, чтобы вместить всю эту информацию. Но здесь встает следующая проблема — проблема времени. Для ввода одного слова в ЭВМ, снабженного всеми нужными признаками, смысловыми и грамматическими, необходим день работы одного исполнителя. Значит, чтобы ввести миллион слов в ЭВМ, нужен один миллион человеко-дней или тридцать лет работы коллектива, состоящего из ста человек.

Эта проблема, в свою очередь, порождает проблему, которую называют «парадоксом Ахиллеса и черепахи». Парадокс этот известен со времени античности: может ли быстроногий Ахиллес догнать черепаху, если в каждый конкретный момент времени черепаха также продвигается вперед? Если следовать правилам логики, Ахиллесу ее никогда не догнать! Не получится ли сходная ситуация и при машинном переводе: информация, заложенная в ЭВМ, будет устаревать к тому моменту, когда начнется практическое применение «электронного мозга»? Ведь слова, как известно, меняются со временем. И никакому Ахиллесу-роботу не догнать наш неспешный, но постоянно изменяющийся язык-черепаху...

Более того, слова не являются, строго говоря, отдельными изолированными единицами языка. Смысл их связан со смыслом других слов, все слова как бы прошиты незримыми нитями ассоциаций. Причем границы между этими связями нечетки, они как бы размыты. Приведем такой пример. В нашем языке есть группа прилагательных, относящихся к возрасту человека: детский, отроческий, юношеский, молодой, средних лет, пожилой, старый. Но попробуйте-ка четко и однозначно распределить смысл этих слов по строгой шкале лет, и вы убедитесь сами, насколько размыты границы между детским и отроческим, отроческим и юношеским, юношеским и молодым, молодым и средних лет, средних лет и пожилым, пожилым и старым у разных людей, в зависимости от их собственного возраста (вспомните журналиста из «Двенадцати стульев», считавшего стариком всякого, чей возраст превысил двадцать лет!).

Теория нечетких множеств — так называется новая область математики, одна из самых интересных и перспективных, которая родилась в связи с описанием языка. Ее аппарат начинает применяться и для машинного

перевода, и для информационного поиска. А надежной точкой опоры становятся те холодные числа, данные статистики, о которых рассказывал наш первый очерк. И если первые опыты по машинному переводу были в полном смысле слова опытами, то в наши дни начинается промышленная эксплуатация МП.

### ДИАЛОГ ПРОДОЛЖАЕТСЯ...

Конечно, никто из ученых не пытается сейчас дать совершенный перевод с помощью машины, подобный переводу человеческому. И переводится не любой текст, а текст по какой-либо узкой специальности, например химии полимеров или низкотемпературной плазме. Но ведь именно такой подстрочный, так сказать, рабочий перевод текстов по специальности и необходим в первую очередь ученым, инженерам, техникам, работающим в той же химии полимеров или в области низкотемпературных плазм.

Первая в нашей стране промышленная эксплуатация МП была осуществлена в 1976 году в Чимкентском пединституте по заказу Института химии Академии наук Казахской ССР. ЭВМ системы «Минск» переводит английский текст длиной в тысячу слов, то есть около трех страниц, за двадцать — двадцать пять минут. И это, учитывая ввод текста в машину и вывод его из машины в виде отпечатанного русскими буквами перевода!

В конце 1977 года в Алмаатинском энергетическом институте проходила VII Всесоюзная конференция по генераторам низкотемпературной плазмы. Участвовали в ней и гости из других стран. Делегатам и гостям были вручены три компактные книжечки в красном переплете: англо-русский и русско-английский, французско-русский и русско-французский, немецко-русский и русско-немецкий словари-минимумы. Предназначены они для чтения научных текстов по плазме, а также для разговора с зарубежными коллегами. Автором же этих словарей была... вычислительная машина и, естественно, коллектив программистов и лингвистов, составивших машинную программу. За три месяца ЭВМ проделала работу, на которую потребовалось бы несколько лет труда сектора научного учреждения.

Перед нами широкий лист бумаги. Вверху напечатан запрос: «Прошу определить общую тему докумен-

та». Далее следует французский текст статьи по технологии окраски. Внизу дается ответ на русском языке: «Документ относится к теме «электрофорез». Затем снова запрос: «Прошу дать машинный реферат документа». На него тут же дан развернутый ответ.

«В этом диалоге, казалось бы, нет ничего удивительного,— пишет профессор Р. Г. Пиотровский, руководитель группы «Статистика речи». — Именно таким образом сотни референтов-переводчиков ведут беседу со своими заказчиками. Однако приведенный текст представляет собой человеко-машинный диалог, в котором вопросы формулировал человек, а ответы на правильном русском языке без всякой подсказки выдавала машина. Учителями компьютера были языковеды, химики и математики из лаборатории инженерной лингвистики Ленинградского педагогического института имени А. И. Герцена и Кишиневского политехнического института».

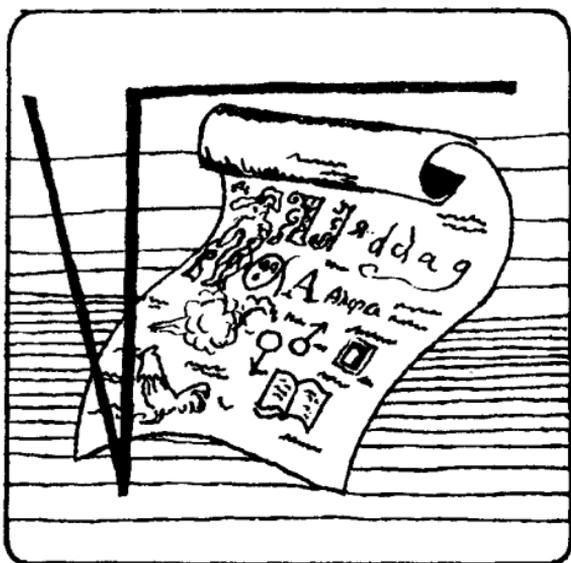
Лишь совсем недавно человеческий мозг создал «мозг» электронный, вычислительные машины. На первых порах диалог человека с ЭВМ казался фантастикой. Затем настала пора радужных надежд и горьких разочарований. Но работа продолжается, продолжают поиски подходов и методов, с помощью которых вычислительные машины включатся в человеческое общество и «электронный мозг» станет не просто чудо-счетчиком, но и надежным и верным помощником в тысяче других дел. Путь к этому — обучить машины человеческому языку.

Диалог человека и работа только-только начинается! Мы же закончим наш рассказ о МП — машинном переводе, о поисках ЯП — языка-посредника и о создании ИЯ — информационных языков словами Норберта Винера, которого называют «отцом кибернетики»: «Отдайте же человеку — человеческое, а вычислительной машине — машинное. В этом и должна, по-видимому, заключаться разумная линия поведения при организации совместных действий людей и машин. Линия эта в равной мере далека и от устремлений машинопочклонников, и от воззрений тех, кто во всяком использовании механических помощников в умственной деятельности усматривает кощунство и принижение человека».

*Передача смысла, значения — вот цель нашего общения, главная задача человеческого языка и, стало быть, главная задача науки об этом языке. Можно ли описывать это значение на языке точных формул и чисел?*

*О том, как математическая лингвистика штурмует «святая святых» языка, расскажет очерк*

## В ПОИСКАХ ЗНАЧЕНИЯ



### ПУТЬ К СЕМАНТИКЕ

Не так давно в сектор структурной и прикладной лингвистики Института языкознания Академии наук пришел человек с рюкзаком. В рюкзаке лежали словари: индонезийского, английского, русского, древнегреческого и других языков.

Человек с рюкзаком объявил: он сделал важное открытие. Чтобы определить степень развитости того или иного языка, надо мерять слова. В буквальном смысле мерять — линейкою. Оказывается, древнегреческие слова на пять миллиметров длиннее индонезийских. Стало быть, и сам язык эллинов более развит...

Конечно, это курьез. Но еще каких-нибудь два десятка лет назад даже у многих серьезных ученых представление о математической лингвистике принципиально не

отличалось от представлений человека с рюкзаком. Помните, как великий комбинатор Остап Бендер, продавая астролябию, напутствовал покупателя словами: «Сама меряет — было бы что мерять». Математической лингвистике отводилась такая же роль: с ее помощью, дескать, все можно в языке измерить, были бы языки!

Но сам термин «математическая лингвистика» не совсем точен. «Нельзя сравнивать термин «математическая лингвистика» с аналогичным термином «математическая физика», — писал профессор Р. Л. Добрушин в годы, когда математическая лингвистика делала свои первые шаги. — Математическая физика — это особый раздел математики, нацеленный на специфические физические приложения; по своим методам он не менее сложен, чем любой другой раздел математики. В лингвистике же речь должна идти о первых шагах применения математики».

В пионерских работах применялся традиционный аппарат теории вероятностей и теории множеств, математической статистики и теории информации. Однако в наши дни начинается создание и своего особого математического аппарата для лингвистики, подобно тому, как он был создан для экономики с ее линейным и динамическим программированием, теорий игр и теорий массового обслуживания. Особые лингвистические проблемы привели к тому, что в математике родилась совершенно новая область — теория формальных грамматик.

В предыдущем очерке мы упоминали теорию нечетких множеств. Она также была создана математиками для решения задач лингвистики. Позже выяснилось, что аппарат этой теории имеет самые различные применения, вплоть до метеорологии. Сейчас теория нечетких множеств — одна из самых перспективных и бурно развивающихся отраслей современной математики (в нашей стране в издательстве «Знание» вышла брошюра основателя этой теории, Л. Заде, «Основы нового подхода к анализу сложных систем и процессов принятия решений» и его же монография «Понятие лингвистической переменной и ее применение к принятию приближенных решений» в издательстве «Физматгиз»).

Когда ученые попробовали применить к языку и его правилам строгие и однозначные алгоритмы, выяснилось, что они слишком просты и грубы. Теперь математики разрабатывают так называемые нечеткие алго-

ритмы, опираясь на теорию нечетких множеств. Образцом такого алгоритма может быть поведение слепого, когда он приближается к цели, постепенно уменьшая свои шаги...

Таким образом, в современной лингвистике наряду с традиционным аппаратом математики есть и свои специальные средства. Они применяются для описания диалектов и изменения языка во времени, для грамматики и лексики. Но при всем многообразии идей и направлений, поисков и перспектив можно увидеть общую тенденцию, которая отличает их от более ранних исследований. Тенденцию эту можно охарактеризовать словами, вынесенными в заголовок: поиски значения. Поиски путей, позволяющих проникнуть в сокровенные глубины языка, а вместе с тем — и вообще человеческого мышления, неотторжимого от языка.

На первом этапе развития математической лингвистики ученые принципиально отказывались от анализа смысла, от содержательной интерпретации высказываний, слова и т. д. Их интересовала система «языка вообще», система конкретных языков, абстрактные формулы грамматики, приложимые к этим языкам, структура языка «икс» нерасшифрованного текста, ритмическая «сетка» правил стихосложения.

Пионерская работа такого плана была сделана... около двух с половиной тысяч лет назад. Тогда великий индийский ученый Панини дал блестящий образец строго формализованного описания санскрита. С появлением быстродействующих электронных вычислительных машин как из рога изобилия посыпались работы, где давались в терминах математики формализованные описания структур различных языков. Однако вскоре стало ясно, что описание грамматики языка или его частотных характеристик, несмотря на всю свою математическую точность, недостаточно еще для решения не только основных задач языкознания, но и для решения частных и даже сугубо практических задач.

Например, для машинного перевода и автоматического реферирования текста необходимо обращаться к смыслу, к значению (вспомните язык смысловых множителей в очерке МП, ЯП, ИЯ). Для расшифровки древних текстов — таких, как кохау ронго-ронго острова Пасхи, где грамматические показатели отсутствуют почти полностью, — методы позиционной статистики, опи-

рающейся на выявление грамматики языка «икс», непригодны. Мало чем полезна позиционная статистика и при анализе ритмики стиха, при котором неизбежно следует обращение к значению, к интерпретации того или иного «звукообраза», создаваемого этим ритмом.

Профессор Роман Осипович Якобсон привел такое красочное сравнение во время своей лекции в Московском институте иностранных языков имени Мориса Тореза. В лингвистике, игнорируя значение, мы даем формализованное описание языка, подобное тому, какое дает человек, описывая поведение курицы, предварительно отрубив ей голову. Описание это будет точным и непротиворечивым... Но разве из этого следует, что оно описывает поведение курицы с головой во всех его аспектах?

Отсюда, конечно, не следует, что математическая лингвистика — тупиковый путь, что ее итогом будет не кибернетическая формализация, а пустой формализм. Нет, в наши дни от формализации чисто «внешнего» в языке лингвистика переходит к формализации «внутреннего», от грамматики переходит к семантике, к анализу смысла.

### «КУЗДРА», «БОКР», «БОКРЕНОК», «ГЛОКАЯ»

Различие между ранним и нынешним этапами математической лингвистики очень хорошо показывает такой пример. Академик Л. В. Щерба давал своим ученикам для анализа, казалось бы, заумную фразу: *глокая куздра штеко будланула бокра и кудрячит бокренка*.

Ни в каком словаре русского языка вы не отыщете всех этих слов, хотя грамматическое оформление фразы русское (Щербе принадлежит крылатый афоризм для изучающих иностранные языки: «Лексика — дура, грамматика — молодец!»), перефраз суворовского афоризма о пуле и штыке).

Опираясь на грамматику русского языка, можно многое выяснить в этой фразе, дать ее расшифровку. Слово *куздра* — женского рода, единственного числа. С ним согласовано стоящее впереди слово *глокая* — в роде и числе. Отсюда вывод: слово *куздра* существительное, слово *глокая* — прилагательное к нему.

Обратимся к глаголам. Очевидно, что ими будут слова *будланула* и *кудрячит*. Слово *будланула* согласо-

вано со словом *куздра* в роде (женский) и числе (единственное). Значит, оно будет сказуемым при подлежащем *куздра*. Глагол *будланула* образован, очевидно, от инфинитива *будлануть* и явно дан в прошедшем времени. Другой глагол *кудрячит* столь же явно обозначает настоящее время, единственное число и также согласован с гложкой *куздрой*.

*Бокр* — существительное мужского рода, единственного числа, потому что этого бокра и будланула гложкая *куздра* (слово *бокр* находится в винительном падеже). Но не просто будланула, а *штеко*. Отсюда вывод: слово *штеко* — наречие.

Остается слово *бокренок*. Вывод напрашивается сам собой: это существительное мужского рода, единственного числа, которое, как и бокр, стоит в винительном падеже...

Дадим формальный анализ всей фразы: (кто?) *куздра* (какая *куздра*?) *гложкая* (что сделала?) *будланула* (кого будланула?) *бокра* (как будланула?) *штеко* и (что еще делает *куздра*?) *кудрячит* (кого *кудрячит*?) *бокренка*. Здесь легко найти, что во фразе является подлежащим, сказуемым, определением и т. д. Иными словами, не зная смысла фразы, мы выявляем ее грамматическую структуру.

Все это — своеобразная характеристика первого этапа формализованного изучения, языка, этапа, на котором смысл, значение игнорируются. В настоящее время можно предложить анализ нашей фразы с гложкой *куздрой* уже не абстрактно-грамматический, а семантический, смысловой. Причем опираться мы будем по-прежнему на саму структуру, только уже не внешнюю грамматическую, а внутреннюю, смысловую.

Начнем с глагола *будлануть*. При нем есть прямое дополнение — *бокра*, которое выражено одушевленным существительным (*бокр* имеет окончание на *-а* в винительном падеже; если бы это существительное было неодушевленным, оно бы имело нулевое окончание, сравните склонение слов *тигр* и *бобр*, подобных *бокру*). Отсюда вывод: глагол *будлануть* переходный. Расчленим его на части. *Будл* — основа, *-ануть* — суффикс.

В русском языке к подобного рода глаголам относятся: *давануть*, *долбануть*, *звездануть*, *мазануть*, *рубануть*, *садануть*, *стегануть*, *толкануть*, *щипануть*, *хлестануть* и т. д. Все они имеют однократное значение, выра-

жают энергичное, насильственное воздействие на объект (нечто вроде ударить, но непременно ударить с силой и один раз). Правда, есть одно исключение — глагол *сказануть*, но он никак не подходит для аналогии с *будлануть*: можно *сказануть* на бокра, но нельзя *сказануть* бокра. Значит, глокая куздра энергично и насильственно воздействовала на злосчастного бокра. Затем она стала кудрячить бедного бокренка.

Обратимся к этому глаголу. Кудрячить имеет сходное значение насильственного воздействия на объект. Этот глагол, как и *будлануть*, — переходной, имеет в качестве прямого дополнения одушевленное существительное. В предложении он связан с *будлануть* соединительным союзом *и*. Кудрячить и *будлануть* — однородные члены предложения. Для подобного рода глаголов, выполняющих одинаковые грамматические функции и связанных соединительным союзом *и*, характерно и смысловое согласование.

Попробуйте сами подобрать в пару к глаголам на *-ануть* другой глагол во фразе, подобной нашей глокой куздры, и вы увидите, что вторым глаголом (эквивалентом кудрячить) обязательно должен быть глагол, имеющий подобный же «агрессивный» смысл. Сопоставьте вот эти две фразы: «Он долбанул его и видит его брата»; «Он долбанул его и лупит его брата». Обе фразы стилистически корявые, но первая в этом смысле более корявая, чем вторая — вторая при некоторой снисходительности к стилю не вызывает никаких возражений: в смысловом отношении она абсолютно правильная.

Итак, глаголы *будлануть* и *кудрячить* имеют четко выраженную смысловую направленность.

Проанализируем теперь наречие *штеко*. Что можно сказать о его смысле? Очевидно, что оно характеризует глагол *будлануть*. В его значение входит признак интенсивности, нечто вроде крепко, ловко. Образовано оно, вероятно, от прилагательного *штейкий* (подобно тому, как ловко образовано от ловкий, крепко — от крепкий и т. п.). В силу этого оно не может быть обстоятельством места, времени, цели, причины и т. д., а дает характеристику глаголу *будлануть*. Опять-таки обратимся к глаголам на *-ануть*. Все качественные наречия, относящиеся к ним, непременно выражают признак интенсивности действия. В таком ряду, как *крепко звезданул*,

ловко мазанул, шибко трепанул, равноправным членом встанет и наше *штеко будланул*.

А что можно сказать о бокре и бокренке? Они образуют пару, где общий корень *бокр*. Слово *бокренок* образовано от *бокра* при помощи суффикса *-нок*. И бокр и бокренок — одушевленные существительные мужского рода. Все это заставляет нас сделать вывод, что бокр — животное, самец, а бокренок — его детеныш.

В самом деле: сравните бобр — бобренок, тигр — тигренок, зверь — зверенок, жеребец — жеребенек, кот — котенок, сом — соменок. Сюда прекрасно вписывается пара бокр — бокренок.

Остается *глокая куздра*. То, что прилагательное *глокая* характеризует *куздру*, ясно. Но ничего другого мы о нем сказать не можем. Куздра может быть морская или речная, мохнатая или гладкая, черная или пегая, старая или молодая, мощная или тихая — одним словом, *глокая*. Значение этого слова-уродца может толковаться по-разному, ибо у нас нет для него рамок, подобных тем, что были найдены для глагола *будлануть* или наречия *штеко*. Единственное, что мы можем сказать определенного о прилагательном *глокая* — это то, что оно входит в характеристику живого существа — *куздры*.

Правда, возникает вопрос: почему мы должны считать *куздру* живым существом? Для *бокра* и *бокренка*, как вы помните, определительным признаком их одушевленности послужило окончание *-а* в винительном падеже. Наша *куздра* стоит в падеже именительном, может быть, это не живое существо, а какой-то предмет, орудие, снаряд и т. п.? Нет, ответим мы, если вспомним, что *куздра* *будланула*. Только живое существо может производить такое целенаправленное действие, как *будлание* — это мы доказали рядом глаголов *тряхануть*, *долбануть*, *давануть*, *щипануть* и т. п., обозначающих действие, которое способно осуществить лишь живое существо.

Что мы еще можем сказать о *куздре*? Было бы соблазнительно определить ее пол: если бокр — самец, то, быть может, *куздра* — самка, раз слово это стоит в женском роде? Действительно, большинство названий животных попадает под эту модель: тигр — тигрица, лев — львица, лис — лисица, за исключением некоторых

слов вроде пантеры — она может быть и самкой, и самцом.

Что же получаем мы в итоге? Куздра, некое живое существо, по всей вероятности, самка, интенсивно произвела насильственное действие над другим существом и оказывает воздействие на детеныша этого существа.

«Этот анализ объясняет, почему подавляющему большинству не искушенных в лингвистике носителей русского языка, к которым автор обращался с просьбой дать толкование щербовской фразе, представлялась приблизительно одна и та же картина: самка сильно ударила какого-то самца и наносит удары его детенышу», — пишет Ю. Д. Апресян, советский специалист по структурной семантике.

### НЕОБХОДИМОСТЬ И ВЕРОЯТНОСТЬ

Первый анализ — грамматический — был предложен Л. В. Щербой. Второй, семантический, заимствован нами из книги Ю. Д. Апресяна «Идеи и методы современной структурной лингвистики». Можно даже усилить аргументы и попытаться более точно определить бокра и его детеныша. По всей вероятности, этот бокр должен быть диким, а не домашним животным. Вспомните, что потомство животных, издавна прирученных человеком, в русском языке (и не только русском, но и во многих других языках мира) называется словами, отличными от тех, которыми именуется взрослая особь. У собаки рождаются щенята, а не собачата, у овцы — ягненок, а не овчонок, у коровы — теленок, а не коровенок, а у свиньи — поросенок, а не свиненок. Зато у тигрицы — тигренок, у львицы — львенок, у орла — орленок. Так что бокр, имеющий бокренка, животное, по всей видимости, дикое, а не домашнее.

А теперь зададимся иным вопросом. Интуитивно всем нам, носителям языка, благодаря грамматике и привычным смысловым связям, рисуется примерно одинаковая картина того, что произошло с глокой куздрой академика Щербы. Но обязательно ли именно такое толкование?

Начнем с первого слова *глокая*. Мы определили его как прилагательное к слову *куздра*. Но ведь его можно трактовать и как наречие, и как деепричастие. Судите сами: куздра может быть свирепой, сильной, хищной и

т. д. Однако может быть и некой, эдакой, и т. п. Как известно, мы не во всех случаях деепричастие в начале предложения выделяем запятой. Поэтому слово *глокая* можно трактовать и как деепричастие, типа прыгая, рыская, падая, воя, рыдая.

Мы интуитивно относим существительное *куздра* не просто к разряду живых существ, но именно к животным, а не людям. Однако в русском языке есть целый ряд слов, относящихся не к литературному, а обиходному пласту лексики, которым созвучна наша *куздра*. Вспомните такие слова, как *мыра*, *грымза*, *карга*, *дура* и т. п. Все они относятся к женщинам, а не животным. Более того, интенсивное и целенаправленное действие может осуществлять не само животное, а какой-либо его орган, например *иога* или *лапа*... Да и неодушевленный предмет также может производить такое действие, если он находится в руках человека: почему бы *глокой* *куздре* не быть эквивалентом железной *кочерги* или *мокрой тряпки*?

Слово *штеко* и *Щерба*, и *Апресян*, и все мы интуитивно считаем наречием. Казалось бы, в этом нет сомнения. Однако припомним названия экзотических животных, вроде *дикой собаки динго* или *кенгуру*. Может быть, не *куздра*, а *штеко* является тем самым одушевленным существом, которое так свирепо будлануло *бокра* и *кудрячит бокренка*?

Преподаватель В. А. Шорохов провел интересный эксперимент со своими учениками, студентами Института иностранных языков имени Мориса Тореза. Он подобрал восемь русских фраз, построенных по типу *глокой куздры* (где и *куздра*, и *штеко*, и *бокр* с *бокренком* получали различное смысловое и грамматическое оформление), и предложил студентам определить, какая из этих фраз по смыслу соответствует *щербовской*. Вот эти фразы: «*Закусочная турка нагло обманула казака и дурячит казачонка*»; «*Ушлая экономка только обманула повара и морочит поваренка*»; «*Дикая собака динго трепанула волка и увечит волчонка*»; «*Описывая мальчика бегло черканула сестра и строчит сестренка*»; «*Слушая брата сильно психанула сестра и молчит сестренка*»; «*Выискивая зерна проворно гребанула ворона и кормит вороненка*», «*Завидевшая обидчика высоко маханула обезьяна и манит обезьяненка*»; «*Наглая телка крепко долбанула пастуха и увечит пастушонка*».

Помимо того, изъяв все корни и оставив лишь одни грамматические окончания, Шорохов составил абстрактную модель щербовской фразы: «А-ая Б-а В-о Г-анула Д-а и Е-ит Д-енка», после чего предложил студентам ее интерпретировать. И тут будущие лингвисты, профессиональные переводчики стали в тупик. Насколько однозначно и легко они трактовали глокую куздру, настолько тяжело им было дать смысловую интерпретацию модели, хотя, казалось бы, сделать это еще легче, чем с самой глокой куздрой: ведь каркас есть — наполняй его чем угодно.

В чем тут дело? А видимо, в том, что условный глагол *будлануть* вызывает у нас ассоциации с реальным глаголом *бодануть*. Мы воспринимаем щербовское слово, имея уже подсознательную трактовку этого глагола как действия, осуществленного живым существом по отношению к другому живому существу. Иными словами, мы заранее запрограммированы на это. И когда начинаем вроде бы сугубо академический, формальный, грамматический анализ фразы, на самом деле уже имеем в голове ее смысловую трактовку. Мы дешифровали фразу, поняли ее, а потом уже разобрали по косточкам с помощью грамматики. Но ведь как раз эта самая грамматика показывает, что глокую куздру можно понимать совсем иначе. Значит, все дело не в грамматике, а в семантике. Не абстрактная геометрия грамматики, а конкретное значение, носителем которого является мозг человека, определяет трактовку придуманной глокой куздры!

...Да простит читатель автора за столь долгий разбор бессмысленной, казалось бы, фразы. Но он был нужен для того, чтобы всякий человек, владеющий русским языком, хотя и не постигший премудростей структурной семантики, смог понять и прочувствовать всю сложность и неоднозначность вопросов, на которые отвечать — и со всей строгостью науки! — предстоит той же самой структурной семантике.

Ибо, как превосходно сказал один из основоположников этой науки, польский ученый Альфред Тарский, «в обычной речи не существует фразы, имеющей точно определенный смысл. Едва ли можно было бы найти двух человек, которые употребляли бы слово в одинаковом значении, и даже в речи одного человека значение одного и того же слова меняется в различные

периоды жизни. Сверх того, значение слов повседневного языка обычно очень сложно; оно зависит не только от внешней формы слова, но также и от обстоятельств, при которых оно высказано, а иногда и от субъективно-психологических факторов».

Фраза, придуманная академиком Щербой, считается образцом, классикой формального анализа, то есть анализа, построенного лишь на законах грамматики. Но мы с вами убедились, что решающая роль здесь принадлежит ассоциациям, связанным у любого человека, владеющего русским языком, с глаголами типа бодануть.

### СТАТИСТИКА АССОЦИИИ

Об ассоциациях, связях, запечатленных в нашем мозгу, писал еще Аристотель. В конце прошлого — начале нашего века были проведены первые эксперименты по установлению словесных ассоциаций. Схема экспериментов такая: человеку предлагают слово и требуют, чтобы он тотчас же назвал «первое пришедшее в голову» другое слово.

Казалось бы, мало ли что может прийти человеку в голову? Ведь ответ его зависит от образования, воспитания, культуры, возраста, пола, характера, настроения, наконец каких-либо текущих впечатлений, например, прочитанной недавно книги или увиденного спектакля. И все-таки, как показали исследования, наиболее употребительные слова рождают и определенные стандартные ассоциации. Причем если смысл слов в разных языках может быть один и тот же, то ассоциации, этим словом вызываемые, бывают порой различны.

Возьмем слово *орел*. Более половины участвовавших в эксперименте американцев на это слово-стимул отвечало *птица*.

А вот у казахов ассоциации совсем иные. Лишь четырнадцать процентов опрошенных дали ответ *птица*. Зато почти шестьдесят процентов казахов отвечали на *орел* словами *беркут*, *лисица* или *заяц*. Почему? Да потому, что орел-беркут, испокон веков использовался казахами для охоты на лисицу, зайца и других животных.

В 1910 году вышел в свет первый специальный словарь ассоциаций, составленный Грейс Кент и А. Розановым. В него вошли ответы тысячи испытуемых, данные

на сто употребительных слов английского языка. Вслед за этим словарем появились словари ассоциаций французского, немецкого и других языков.

Совсем недавно мы получили «Словарь ассоциативных норм русского языка», включающий в себя полтысячи слов и многочисленные слова-ассоциации, которые порождало то или иное слово у нескольких сотен испытуемых. В словаре приводится и статистика этих ассоциаций, начиная с самых частых ответов и кончая единичными. Понятно, что эти числовые данные говорят о том, насколько прочна и стандартна связь того или иного слова русского языка с другим.

Возьмем, например, слово *дядя*. Более трети всех испытуемых ответили *тетя*. Это, так сказать, самая обычная ассоциация на *дядю*. Весьма частыми оказались ответы: *мой, Ваня, родной, Степа, добрый, чужой, родственник, самых честных правил, хороший*. Они порождены различными причинами: сочетаемостью слов, стандартными оборотами, литературными откликами («Дядя Ваня» — пьеса Чехова; популярная строка из «Евгения Онегина» «Мой дядя самых честных правил»; «Дядя Степа» — персонаж Михалкова). А вот единичные ответы вроде *низкого роста, нудный, сантехник, Олег, футбол, батя, безумный, ирония, Гоша, река, редко вижу, прислать, полицейский* и т. п. вызваны причинами чисто субъективными. У кого-то дядя низкого роста, у кого-то дядю зовут Олегом или Гошей, у кого-то дядя работает сантехником, у кого-то дядя — нудный человек и т. п.

Еще в начале пятидесятих годов нашего века была предложена методика «вычисления осмысленности» слова на основании числа ассоциаций, которые оно вызывает. Наиболее интересные результаты она дает в том случае, если мы сознательно ограничим свою задачу. Например, будем требовать не ассоциации вообще (они, как вы сами убедились на примере слова *дядя*, могут быть самыми неожиданными), а определенного типа ответы. Например, называя слово, мы просим привести уже не «первое пришедшее в голову» слово, а противоположное названному по смыслу.

В нашей стране был проведен интересный эксперимент на более чем пятистах испытуемых. Им предлагалось слово, испытуемые называли в ответ противопоставление. Затем эти ответы обрабатывались статисти-

чески, и мы получали возможность выразить в числах степень связи значений слов.

В ответ на слово *слепой* сто процентов испытуемых дали ответ *зрячий* (смысловый индекс противопоставления, таким образом, будет равен единице). У слова *темный* этот индекс оказался равен 0,94 (почти все испытуемые ответили *светлый*), у слова *пустой* — индекс 0,82 (ответом было *полный*) и т. д.

Но это, так сказать, один полюс противопоставления. На другом полюсе находятся слова, для которых нет столь однозначных ответов. Слово *сердитый* получило четыре примерно равных ответа: *веселый, добрый, радостный, добродушный* (а вдобавок к ним более редкие: *беззлобный, ласковый, приветливый, доверчивый*). Между этими двумя полюсами оказались слова, имеющие большой набор индексов, но при этом один из них доминировал. Например, слово *защита* имело частый ответ *нападение* (индекс 0,72), но помимо него, назывались слова *атака, беззащитность, предательство, бессилие, доступ, брешь, удар, незащищенность, бегство, вторжение, прорыв, захват, угроза*. Причем у всех этих слов индексы были очень малы, от шести сотых до одной сотой, то есть называла их ничтожная часть испытуемых.

Предлагались и другие методики измерения значений на основе ассоциаций. Например, вычисление степени близости слов-стимулов. Стандартная ассоциация на английское слово *girl* (девушка, девочка) — слово *boy* (мальчик), ее дает более семидесяти процентов испытуемых. Кроме того, были ответы со значением *женщина, друг, молодая, платье, хорошенькая, милая* и т. д. Стандартная ассоциация на слово *boy* — слово *girl*. Смысловая близость слов *boy* и *girl* вычисляется по формуле и равна 0,749; не следует думать, однако, что ассоциации двух слов взаимны; в ответ на слово *table* восемьдесят пять процентов ответило *chair*, а на *chair* ответ *table* не дала и половина всех испытуемых.

Советская исследовательница А. П. Клименко предложила не косвенный, а прямой метод измерения смысловых связей двух слов. Испытуемым предлагается оценить, исходя из шкалы в десять баллов, близость смысла двух слов. Данные, полученные в результате опроса многих людей, обрабатываются, и получается некая средняя величина. Вот какие числа были получены в экспериментах. Смысловая близость слов *стол* и *год*,

стол и время равна почти нулю (индексы 0,4 и 0,6). Между *время* и *погода* она равна четырем баллам, между *год* и *месяц* — 7,4, между *снег* и *погода* — 7,1. Смысловая же близость между словами *время* и *час* приближается к максимальной, она равна 9,1 балла.

Любопытно сравнить эти данные с результатами, полученными по словарю ассоциаций. Наиболее частый ответ на слово *час* — слово *время* (более двадцати процентов всех ответов). Наиболее частый ответ на слово *время* — *час* (примерно десять процентов всех ответов). А самым частым ответом на слово *год* слово *месяц*, его дает приблизительно пятнадцать процентов всех носителей русского языка. Значит, как бы ни различалась методика измерения значений, она приводит к сходным результатам. И это говорит о том, что в своих поисках значения исследователи стоят на верном пути.

### «ИЗМЕРЕНИЕ ЗНАЧЕНИЙ»

«Измерение значений» — так называлась монография американских лингвистов Осгуда, Танненбаума и Сучи. Она вызвала широкие отклики во всем мире. В ней рассказывалось об огромной работе, проведенной Осгудом и его сотрудниками по определению смысловых связей слов или, как говорит само название монографии, по измерению значений.

В любом языке мира есть полярные, противоположные по смыслу термины и слова: *север* — *юг*, *холод* — *жара*, *большое* — *маленькое*, *красивое* — *уродливое*, *старое* — *новое* и т. д.

Быть может, смысл слова удастся определить как некоторую точку, что находится на этих полярных словах-шкалах? Нельзя ли взять в качестве измерений, неких координат ряд подобного рода полярных, противоположных по смыслу слов? А все остальные слова расположить по этим координатам?

Возьмем, к примеру, пару *хороший* — *плохой*. Предмет или явление может быть или плохим, или хорошим. Однако в языке есть более тонкое деление: мы говорим — и оцениваем — явление или предмет как очень хорошее, очень плохое, не плохое и не хорошее (то, что называется ни рыба, ни мясо). Есть и, так сказать, средние оценки, те, что выражаются словами вроде *неплохой*, *ничего себе* и т. п.

Понятие «не плохой и не хороший» нейтрально, безразлично, и Осгуд предложил обозначить его через нуль. Понятие «немного хороший» будет тогда оценено в плюс единицу, а «немного плохой» — в минус единицу. «Вполне хороший» оценивается в плюс два, «вполне плохой» — в минус два. Наконец, крайние точки шкалы — «очень плохой» и «очень хороший» — обозначены Осгудом как плюс три и минус три. Таким образом, шкала значений *хороший* — *плохой* получила семь делений, семь возможных числовых оценок, от минус трех до плюс трех.

Точно так же на семь различных градаций разделено было расстояние и между другими парами противоположных по смыслу прилагательных: *счастливый* — *печальный*, *быстрый* — *медленный*, *ласковый* — *жестокий*, *гладкий* — *шероховатый*, *сложный* — *простой*, *длинный* — *короткий*... Записать эти пары прилагательных можно в виде таблицы. Разграфлена она будет на семь частей, семь оценок, которые может иметь то или иное слово, попадающее в смысловое пространство. А пространство это будет задаваться нашими прилагательными, всеми парами типа *хороший* — *плохой*, *длинный* — *короткий*.

Изобразим это наглядно. Вот перед нами таблица смыслового пространства, которое задается четырьмя прилагательными:

<i>хороший</i>	+3	+2	+1	0	-1	-2	-3	<i>плохой</i>
<i>счастливый</i>	+3	+2	+1	0	-1	-2	-3	<i>печальный</i>
<i>твердый</i>	+3	+2	+1	0	-1	-2	-3	<i>мягкий</i>
<i>медленный</i>	+3	+2	+1	0	-1	-2	-3	<i>быстрый</i>

Вы уже догадались, видимо: в таблице вместо чисел (как с плюсом, так и с минусом) проставляются слова, указывающие на убывание какого-то качества по направлению к нулю и на возрастание противоположного качества от нуля.

Вот по такой методике и оцениваются слова, обращаясь при этом к носителям значений, то есть к людям, говорящим на данном языке. Например, в какую клетку поместим мы значение слова *змея*? Вероятнее всего, на шкале *хороший* — *плохой* она получит оценку в минус три, ибо для многих слово *змея* ассоциируется с чем-то плохим, коварным, неприятным. А слово *отец* может получить оценку плюс три (очень хороший). Правда, любители змеиного племени могут оценить зна-

чение слова змея и в плюс три, а обиженные дети оценят отца в минус три как очень плохого.

Чтобы добиться объективных результатов, чтобы выявить, каково же значение того или иного слова в сознании носителей языка, надо опросить как можно больше людей, самых разных возрастов и профессий, так же как поступают при составлении словарей ассоциаций. Только тогда, сопоставив сотни ответов, можно вывести среднюю величину значения интересующего нас слова.

Например, один испытуемый оценил значение слова *отец* как вполне хорошее (то есть дал ему оценку плюс два), немного счастливое (плюс один), вполне твердое (плюс два) и немного быстрое (минус единица). Другой дал оценки соответственно плюс три, плюс три, нуль, плюс один. Третий — новый вариант оценок. Чтобы данные были объективными, Осгуд и его сотрудники привлекли к участию в опытах по измерению значений сто человек. По результатам их опроса и выводилась средняя величина оценки того или иного слова.

Но почему, собственно говоря, взято только четыре пары прилагательных, а не сорок четыре или сто четыре? И почему взяты пары *хороший — плохой*, а не, скажем, *честный — бесчестный*? Сколько надо взять пар, что задают координаты значения слова? Все прилагательные, противоположные по смыслу друг другу? Или же многие пары дадут одинаковые результаты и их можно объединить (например, *справедливый — несправедливый* и *честный — бесчестный*)?

Каждую пару прилагательных можно сравнить с измерением. Тогда вопрос сформулируется так: «Сколько измерений нужно для значения слов, сколько измерений имеет смысловое пространство?»

На первом этапе Осгуд выделил все пары прилагательных английского языка, имеющие противоположный смысл. Таких пар оказалось очень много, около трех сотен. Вряд ли смысловое пространство имеет столько измерений. Да и человек, определяя значение слова на трехстах шкалах, естественно, устанет и будет давать неточные и неверные ответы. Вот почему для первого эксперимента выделено было пятьдесят пар. И сто испытуемых размещали значение того или иного слова уже не по четырем парам, как в нашем примере, а по пятидесяти. И не одно слово *отец*, а два десятка разных слов.

Сто испытуемых, двадцать слов, пятьдесят пар... Итого сто тысяч суждений. Этот своеобразный «куб данных» был подвергнут анализу с помощью электронной вычислительной машины: не находятся ли в прямой связи многие эти пары, нельзя ли свести полсотни пар всего лишь к нескольким?

Уже первые опыты показали, что шкалы *хороший — плохой, справедливый — несправедливый, ласковый — жестокий, честный — бесчестный* и т. п. связаны между собой и дают примерно одни и те же оценки. Зато они совершенно не зависят от таких шкал, как *длинный — короткий, мокрый — сухой, медленный — быстрый* и т. д.

Так сколько же подлинных измерений имеет смысловое пространство?

Математический анализ, проделанный ЭВМ, показал, что оно состоит всего-навсего из четырех измерений.

Первое — **оценочное**. Сюда входят такие пары, как *хороший — плохой, ласковый — жестокий, красивый — безобразный, чистый — грязный* и тому подобные оценочные пары прилагательных.

Второе измерение смыслового пространства — **сила**. Его образуют пары прилагательных *большой — маленький, тяжелый — легкий, сильный — слабый, толстый — тонкий* и т. п. Правда, иногда это измерение бывает связано с оценочным. Например, в таких парах, как *храбрый — трусливый, твердый — мягкий, громкий — тихий*.

Третье измерение можно назвать **ориентированной активностью**. Его образуют пары прилагательных *быстрый — медленный, жаркий — холодный, активный — пассивный* и другие, им подобные. В редких случаях это измерение также может переплетаться с оценочным, например, в парах *молодой — старый, бешеный — спокойный, напряженный — расслабленный*.

Наконец около двух процентов результатов не подходили ни под измерение силы, ни под измерение оценки, ни под измерение активности. Они как бы выпадали из трехмерного пространства измерения значений, образуя четвертое измерение.

Может быть, если анализу подвергнуть не полсотни, а больше пар прилагательных, число этих измерений возрастет? Осгуд и его сотрудники увеличили число пар, доведя его до семидесяти шести. Однако это не увели-

чило число измерений смыслового пространства. А так как четвертое измерение малосущественно, то Осгуд решил оперировать с трехмерным пространством смысла, состоящим из оценки, силы и активности.

Правильность этого решения подтвердила заключительная серия экспериментов. Здесь методика была несколько иной. Сорока испытуемым предъявлялась сначала «основная» пара. Затем они были должны отметить в словах другой пары прилагательных то слово, которое кажется им более близким к этой «основной». Например, за основу бралась пара *тупой — острый*. К какому из слов ближе пара *расслабленный — напряженный*? В большинстве случаев к слову *острый* называлось слово *напряженный*, а к слову *тупой* — *расслабленный*. Такое же согласие было достигнуто и в парах *прямой — кривой* и *благородный — скотский* и т. п. Таким образом выявлялась зависимость пар друг от друга, они находились в одном и том же измерении смыслового пространства.

Зато при сопоставлении таких пар, как *свежий — засохший* и *длинный — короткий* одинаково часто к слову *свежий* назывались и слово *длинный*, и слово *короткий*. По всей видимости, эти пары находились в разных измерениях смыслового пространства.

Перебрав множество десятков пар, Осгуд пришел к выводу, что смысловое пространство состоит из трех измерений. **Оценка, сила, активность** — вот три кита, на которых стоит смысл слов.

## КООРДИНАТЫ СМЫСЛА

Итак, смысл любого слова может быть выражен точкой в смысловом пространстве. Значение слова будет выражаться в виде набора чисел, обозначающих ту или иную величину оценки, силы и активности этого слова. По этим трем шкалам и стали распределяться значения слов в экспериментах Осгуда. Система же оценок осталась прежняя, от плюс трех до минус трех баллов. И каковы бы ни были личные вкусы и субъективные оценки опрашиваемых, но если их достаточно много, то можно получить координаты смысла того или иного слова с достаточной точностью (так же как мы получаем достоверный набор ассоциаций, связанных со словом, опросив достаточно большое число лиц).

Осгуд измерял значение имен существительных. Работы его последователей показали, что таким же образом могут быть измерены значения глаголов, наречий и даже предлогов. Все эти части речи охватываются тремя измерениями смыслового пространства! Так, например, почти все испытуемые определяли глагол *есть* как нечто *доброе, своевременное, вкусное, округлое, сильное, обычное, красивое*.

Результаты измерения значений можно выразить в геометрической форме. Значение слова будет точкой, координаты которой заданы тремя смысловыми измерениями: силой, оценкой, активностью. В этом трехмерном пространстве можно будет разместить значения всех измеренных слов. А затем, используя элементарную формулу, подсчитывать расстояние между тем или иным словом, вернее, между смыслами слов.

Так, Осгуд определил, что расстояние между значением выражений *бутоны белых роз* и *зыбучий песок* равно 8,5; между словами *герой* и *мужественность* — 1,4, а между *героем* и *бутонами белых роз* — 8,5. Вполне понятно, что чем дальше значения слов, тем больше между ними будет расстояние в системе координат смысла. Значения слов *герой* и *мужественность* близки, в то время как *бутоны белых роз* одинаково далеко отстоят от значений слов *герой* и *зыбучий песок*.

Вычислив расстояние между всеми парами значений, можно построить некую карту смысла, небольшой кусочек, отражающий смысловую географию нашего языка, структуры значений его слов. В Соединенных Штатах был выпущен специальный «Атлас смысловых профилей» для значений трехсот шестидесяти различных слов. На его страницах приводились значения этих слов в виде своеобразных карт. Значение каждого слова на карте определялось точкой, координаты которой заданы двумя измерениями: оценкой (хороший — плохой) и силой (слабый — сильный).

Вполне понятно, что на страницах «Атласа» отражена была «география смысла» одного лишь английского языка. Ведь Осгуд работал с людьми, для которых этот язык был родным, и все его данные касались значений слов именно английского языка. Насколько близка «география смысла» в других языках мира? Отличается ли она от английской? Или же координаты смысла всеобщие, подобно нашему трехмерному пространству? И как

на «карте смысла» располагаются значения слов различных языков мира — не так ли, как любая точка земной поверхности, имеющая свою определенную долготу и широту?

Прежде чем ответить на этот вопрос, необходимо решить другой, принципиальный. Достаточно ли в этом случае трех измерений смысла, то есть оценки, силы и активности? Да, для ряда языков мира, как показали опыты, хватает трех измерений смысла, чтобы дать координаты значений слов. Но далеко не для всех языков. Например, в греческом языке нет измерений силы и активности. Зато есть измерение, которого нет в английском.

Еще больше отличался от английского своими измерениями смысла язык одного из индейских племен Северной Америки — навахо. Правда, и в нем было одно из «английских» измерений — оценочное. И, как показали дальнейшие исследования, именно это оценочное измерение является всеобщим для всех языков, подвергшихся анализу: английского, финского, греческого, японского, навахо и многих других. Вполне вероятно, что оно есть во всех языках мира.

Но стоит только перейти к размещениям значений внутри этого всеобщего оценочного измерения, как картина меняется от языка к языку. Мы уже признали, что для европейца змея — это нечно опасное, отвратительное, враждебное. Естественно, что он оценит ее баллом минус три, как «очень плохое». Для аборигенов Австралии змея — это пища, дичь, мясо, и они скорее всего оценят значение этого слова баллом плюс три. Для европейцев собака — друг человека. Значение этого слова мы оценим положительным баллом. А вот на арабском Востоке собака считается нечистым и грязным животным. Наверняка значение этого слова получит отрицательный балл, и учебники арабского языка рекомендуют, говоря — собака, предварительно извиниться перед собеседником, как будто вы произносите нецензурное слово.

Даже в языках, где есть одинаковое число смысловых измерений и эти измерения совпадают, координаты смыслов отдельных слов различны. Например, расстояние между значениями слов *гнев* и *сочувствие* в норвежском языке меньше, чем в английском. С другой стороны, значения слов *горе* и *вина* в английском языке

близки, но в норвежском не связаны друг с другом. А вот какие любопытные результаты показало сравнение значений слов, связанных с культурой и обществом.

«Прогресс есть нечто хорошее — сильное — активное для всех народов, кроме финнов, для которых это нечто пассивное; подобным же образом будущее есть нечто хорошее — сильное — активное для всех народов, кроме финнов, для которых оно хорошее, но слабое и пассивное; труд есть нечто хорошее — сильное — активное для всех народов, кроме фламандцев, для которых это нечто плохое — сильное — пассивное», — пишет Осгуд. И приводит еще несколько характерных примеров.

*Друг* и *мужчина* для американцев и японцев есть нечто хорошее — сильное — активное, однако пассивное для фламандцев и слабое для финнов. Только для американцев со значением слова *полисмен* связано нечто хорошее — сильное — активное. Для фламандцев оно плохое — сильное и пассивное, а для японцев — плохое — сильное — активное. Понятия *поражение*, *бой*, *вор*, *преступление*, *опасность* для американцев есть нечто плохое — сильное — активное, так же как для бельгийцев и финнов. Зато для японцев эти же слова нечто плохое — слабое — пассивное, а *бой* — нечто хорошее — слабое — активное!

Комментарии, как говорится, излишни. Методы измерения значений лишний раз подтверждают, что язык создан обществом и для общества. Разные коллективы, носители разных языков по-разному оценивают значения слов, связанных с определенными культурными и общественными реалиями.

### СДВИГ ЗНАЧЕНИЯ

Если значение отдельных слов можно измерить, то нельзя ли выражать в числах и сочетания слов, их смыслов? Ведь говорим мы не отдельными словами, а предложениями...

Казалось бы, на первый взгляд сделать это очень просто: надо суммировать значения отдельных слов, входящих в словосочетание. Однако такое элементарное сложение ничего не даст. Значения слов воздействуют друг на друга, они не являются некими кубиками, которые можно механически складывать и переставлять. Смысл сочетания слов *исход* и *поиск* будет противопо-

ложным, в зависимости от того, скажем ли мы *исход* *поиска* или же *поиск* *исхода*. А ведь слагаются эти выражения из одних и тех же смыслов слов!

Возьмем три сочетания с прилагательным *жестокий*: *жестокая судьба*, *жестокий ум*, *жестокое окно*. Очевидно, что во всех этих сочетаниях значения слов согласуются по-разному. *Жестокая судьба* — устойчивое сочетание слов, смыслы их прочно спаялись в нашем сознании. *Жестокий ум* звучит несколько необычно, но все-таки приемлемо, например по отношению к холодному рассудительному человеку, бесстрастному аналитику сердечных чувств и т. п. Выражение *жестокое окно* осмыслить можно с большой натяжкой, посчитав это окно тюремным, забранным решеткой, или окном, о которое мы порезали руку...

Во всех трех примерах значение слова *жестокий* меняется в зависимости от слова, с которым оно сочетается. Меняется в зависимости от слова *жестокий* и значение слов *судьба*, *ум*, *окно*.

Нельзя ли, зная координаты смысла каждого слова, взятого в отдельности, измерить меру и степень этих значений? И, что еще более важно, измерить значение всего словосочетания?

Когда два значения соединяются в одно, два слова — в словосочетание, то смыслы их как бы сдвигаются по направлению друг к другу. Степень такого сдвига будет зависеть от меры, от числа, которым выражено значение каждого отдельного слова, входящего в сочетание. Чем дальше друг от друга отстоят они, например по степени оценки, или силы, или активности, тем меньшим будет этот сдвиг.

Слово *судьба* в измерении оценки получило минус один балл (то есть плохое). Слово *жестокий* оценено как очень плохое (минус три балла). Лингвисты, занимающиеся измерением значений, предложили специальную формулу, по которой можно вычислить сдвиг значения слов, входящих в словосочетание. Так, в нашем примере значение слова *жестокий* сдвинулось с минус трех до минус двух с половиной — на полбалла. А значение слова *судьба* с минус единицы сдвинулось до минус двух с половиной, то есть на полтора балла.

Была предложена и формула, по которой можно вычислить значение всего словосочетания. Для примера с

жестокой судьбой это значение будет равно минус двум с половиной баллам.

Проверку формулы провели на очень интересном опыте. Исследователи взяли восемь существительных и восемь прилагательных, а затем составили из них все возможные сочетания. Значения этих шестидесяти четырех сочетаний прилагательного с существительным были вычислены по формуле сдвига значения и по формуле, определяющей значение словосочетания. А затем была проведена проверка на людях — живых носителях значений. Результаты теоретические и практические удивительно близко совпали.

Правда, работы последних лет показали, что эти формулы скорее всего применимы лишь к парам «прилагательное — существительное». Сочетания же иных типов могут следовать другим законам сдвига значения.

Так, например, было доказано, что совсем иной характер носит сдвиг значения, если сочетаются наречие и прилагательное. Если при вычислении значения пары прилагательное плюс существительное мы применяем сложение, то для измерения смысла пары наречие плюс прилагательное нужно переходить к умножению. Тем самым, кстати сказать, объясняется тот факт, что мы воспринимаем как бессмысленные сочетания вроде *немного средний, чрезвычайно обыкновенный* и т. п. Прилагательные *средний* и *обыкновенный* оцениваются нулевым баллом. А умножение на нуль дает нуль: нулевой смысл прилагательного уничтожает смысл и наречия, которое входит с ним в сочетание.

## ОКЕАН ТРУДНОСТЕЙ, ГОРА ПРОБЛЕМ

Мы рассказали об измерении значений, о тех интереснейших результатах, которые были получены. Теперь пора рассказать и обо всех стоящих перед исследователями трудностях, о спорном и нерешенном.

Первая трудность — строгий выбор числа измерений смыслового пространства. В английском их три (хотя и это число оспаривается многими лингвистами, некоторые исследователи вели анализ по восьми измерениям значений). А сколько измерений в других языках? Пять? Два? Десять? Никто не может назвать язык с наименьшим или наибольшим числом измерений смыслового

пространства, никто не может точно расклассифицировать языки мира на двумерные, трехмерные, пятимерные и т. д.

Впрочем, и в английском языке, лучше всех изученном «измерителями значений», также много неясностей со смысловым пространством. Иногда трехмерное пространство смысла как будто теряет одно из своих измерений. Сила и активность прочно сливаются в одно общее измерение — динамизм. Порой же случается, что оно становится и вовсе одномерным, ибо сила и активность поглощаются оценочным измерением. Так что на сегодняшний день даже для английского языка можно с уверенностью назвать лишь одно надежное и выверенное измерение — оценочное.

Но не только в этом заключаются трудности. По самой идее измерения значений эта операция должна быть обратимой: по заданным числовым признакам мы должны всегда разыскать слово, которое эти признаки описывают. На практике это далеко не так. А точнее — почти всегда не так. Попробуйте-ка догадаться, какое понятие описывают признаки «счастливый плюс один», «быстрый плюс два», «твердый плюс один»... Оказывается, речь идет о понятии *отец!*

Или еще один пример. Слово имеет координаты смысла, выраженные в трех его измерениях числами +1,09, —1,85 и +0,77... Что это за слово? Почти наверняка не догадаетесь: это слово *миг*. Между тем, если бы данные по шкалам не были так обобщены, мы могли бы догадаться. Ведь это слово оценивалось в опытах как очень короткое, очень маленькое, очень быстрое, хорошее, скорее простое и т. п. Но при сведении этих оценок в общую пропала их специфика.

Кстати, еще один недостаток измерения значений, который можно показать на примере слова *миг*. Значение слов *момент* и *мгновение* практически одно и то же, что и *миг*. Однако же, по данным «Атласа смысловых профилей», расстояние между ними равно 0,3 — 0,7 — 1,0. Между тем слова, значение которых различается, находятся порой удивительно близко: смысловое расстояние между словами *год* и *век* равно трем десяткам (а синонимы *столетие* и *век* разделены одной и одной десятой).

Вот еще один характерный недостаток: одинаковые расстояния разделяют пары самых различных слов.

Судите сами, какие разнообразнейшие пары оказались разделенными показателями единицы: *час—век, год—сезон, сутки—апрель, июль—час, июль—прошлое, месяц—время, береза—время, воскресенье—завтра, воскресенье—неделя, утро—завтра* и т. п. Интуитивно нам ясно, что на самом деле значения слов *год* и *сезон* гораздо ближе, чем *береза* и *время*. А ведь именно эта интуиция носителя языка — главное в измерении значений, все остальное имеет вспомогательный характер, все эти «кубы данных», формулы и т. д. Значит, ошибаются они, а не интуиция!

Возможно, что ряд ошибок вызван выбором пар прилагательных, через которые пропускалось слово, чье значение измерялось. «Когда на шкалах гладкий — шероховатый, высокий — низкий оцениваются понятия типа столб или гриб, шкалы применяются испытуемыми буквально; когда те же шкалы используются для измерения значений таких слов, как леди, грех, шкалы могут быть поняты только метафорически», — пишет Ю. Д. Апресян, давший вдумчивый и основательный разбор методики измерения значений, разработанной Осгудом. Получается, что приведенные выше пары прилагательных в различных словах измеряют разные значения, в одном случае прямое, буквальное, а во втором — образное, метафорическое. И примеры такого раздвоения значений можно увеличить: так, по-разному на шкале *твердый — мягкий* будут оцениваться слова *камень* и *человек, дерево* и *характер*.

Следует добавить, что слова в английском, как и в любом другом языке мира, очень часто имеют не одно, а несколько значений. В зависимости от того, на каком из этих значений остановится испытуемый, будет находиться и оценка. Например, слово *fire* означает и огонь, и пожар. Очевидно, что оценка несущего тепло и свет огня будет резко отличаться от оценки губительного и опасного пожара.

И все-таки, несмотря на все эти существенные недостатки, работы Осгуда и его сотрудников имеют большое значение, открывая перспективы дальнейших поисков, новых исследований. Ибо, как остроумно заметил один из крупных зарубежных лингвистов У. Вайнрайх, хотя ученые в своих поисках значения и не открыли нового пути в Индию, приобретенный ими опыт навигации может оказаться весьма полезным.

Навигация в «страну значения» началась давно, и пионерами ее были специалисты по математической логике. Фраза «Венера — утренняя звезда» обозначает тот же объект, что и фраза «Венера — вечерняя звезда». Однако очевидно, что значение этих фраз различно... Что же это такое — значение? Этому вопросу посвящают десятки монографий и сотни трудов математические логики, специалисты по теории знаков, философы и математики. И с каждым годом становится все более ясным, что ключ к решению этой проблемы — анализ нашего обычного языка, во всей его внешней простоте и необыкновенной внутренней сложности.

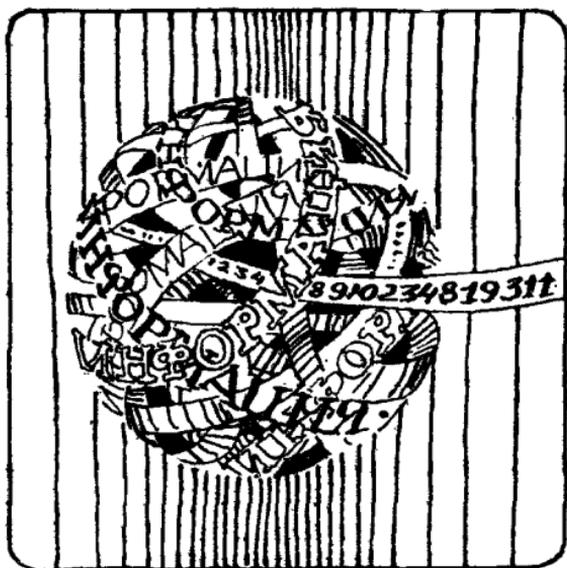
К анализу значения, смысла слов пришла и современная лингвистика. Ибо этого требовала сама логика развития науки о языке. Этому требуют, как вы, вероятно убедились и сами, насущные проблемы машинного перевода, информатики, реферирования литературы с помощью ЭВМ. Словом, к поискам значения привели задачи теории и практики языкознания второй половины двадцатого столетия. И в этих поисках наука о языке идет рука об руку с другими дисциплинами. Лингвист, стремящийся найти путь к измерению значений, обращается к ассоциациям, которые изучает психолог, а математическая статистика делает его выводы достоверными. Анализ значения потребовал создания новых разделов математики вроде теории нечетких множеств и толерантных пространств.

Проблемой смысла занимаются в наши дни не только лингвисты, но и философы, психологи, логики, кибернетики, специалисты по теории знаков — семиотике. Слишком уж сложен и многомерен человеческий язык, главная задача которого — передача смысла.

Недаром же именуют его семантическим кодом.

Математическая теория связи позволяет измерять информацию с помощью точных чисел. О том, как теория информации находит применение в изучении человеческого языка, о сложности этого изучения расскажет очерк

## НАШ УДИВИТЕЛЬНЫЙ КОД



### ФОРМУЛА ШЕННОНА

Для чего мы говорим? Что является целью всякого общения? Зачем в человеческом обществе существуют такие мощные и дальнобойные средства связи, как телевидение, радио, телеграф?

Очевидно, для передачи сведений. Или, говоря другими словами, для передачи информации. Слово *информация* имеет много значений. Но связистам, инженерам, техникам, математикам необходимо одно значение — точное и четкое. «Быстрое усовершенствование техники связи, рост потребностей в передаче информации, «кризис эфира», в котором «не умещается» информация, передаваемая в форме электромагнитных волн, — все это поставило очень остро проблему создания более экономных методов передачи информации», — пишет доктор физико-математических наук Р. Л. Добрушин в статье «Математические методы в лингвистике».

А прежде всего необходимо было ввести точную меру, единицу измерения информации. Еще в 1928 году американский инженер Хартли предложил оценивать количество информации логарифмом числа возможных событий.

Когда мы бросаем вверх монету, ясно, что она может упасть либо гербом, либо решеткой. Если мы бросаем игральный кубик, то неопределенность (или, как говорят математики, энтропия) исхода возрастает. Ведь с одинаковой вероятностью может выпасть любая из граней кубика, желанная шестерка столь же часта, как единица, двойка, тройка и т. д. Понятно, что сообщение о том, какой стороной упала монета, несет меньше информации, чем сообщение о том, сколько очков выпало при бросании кубика. Ибо информация — это то, что снимает неопределенность, то есть, попросту говоря, снимает незнание.

Общепринятой единицей измерения информации считается бит или «да — нет» единица. Слово *бит* происходит от сокращенных английских слов *binary digest* — двоичный разряд, так как для измерения информации в битах берутся не привычные нам со школьной скамьи десятичные логарифмы, а двоичные, основанием которых служит число 2.

Известие о том, что подброшенная в воздух монета упала гербом, принесет нам информацию ровно в один бит. Ведь  $\log_2 2$  («орел» или «решка?») равен 1, то есть одному биту. Известие о том, что выпала игральная карта трефовой, пиковой или другой из четырех мастей, принесет нам информацию в два бита, ибо  $\log_2 4 = 2$ . Сообщение об исходе ситуации, где были возможны (и равновероятны!) восемь вариантов, даст информацию в три бита ( $\log_2 8 = 3$ , или  $2^3 = 8$ , а число битов и есть показатель степени числа два).

Но эта мера удобна и верна лишь при условии, если все наши «выборы» равноправны, имеют одинаковую вероятность появления. И масть игральной карты, и любая грань кубика, и герб или решетка монеты выпадают с равной вероятностью. А как быть, если вероятности не равны?

Хартли понимал, что вероятности исходов влияют на количество информации, которое несет сообщение. Почти невероятному исходу нельзя придавать такое же значение, как и самому правдоподобному. Но он считал,

что различия между этими исходами нельзя выразить в числах. Они определяются психологическими (если речь идет о людях), метеорологическими (если речь идет о погоде) или другими факторами, неподведомственными математике.

Однако в 1948 году американский математик и инженер Клод Шеннон показал, что эта точка зрения ошибочна. Любые факторы — психологические, метеорологические и т. д. — можно учесть, привлекая теорию вероятностей. Он предложил формулу, с помощью которой можно измерять количество информации о событиях, происходящих с разной вероятностью.

Вот эта формула Шеннона:

$$H_1 = - (P_1 \log_2 P_1 + P_2 \log_2 P_2 + \dots + P_n \log_2 P_n).$$

$H_1$  — эта величина неопределенности, которую снимает сообщение, и, значит, мера количества информации (ведь информация уничтожает неопределенность);  $n$  — число «выборов», а  $P_1, P_2, \dots, P_n$  — вероятности появления этих «выборов».

Благодаря этой формуле ученые получили возможность измерять информацию, содержащуюся в кодовых знаках самого различного содержания. Более того, благодаря тому, что мы избираем в качестве «меры» информации логарифмы, мы можем складывать информацию, содержащуюся в каждом кодовом знаке, составляющем сообщение, и таким образом измерить количество информации, содержащееся во всем сообщении.

Действительно, как учит теория вероятностей, вероятность появления двух событий равна произведению вероятностей этих событий. И сумма информации, которую несут кодовые знаки, равна информации всего текста, из этих знаков состоящего. Не будь логарифмов, нам пришлось бы умножать вероятности появления этих знаков. «Логарифмическая» формула Шеннона тем и удобна, что согласно ей информация двух страниц книги — это сумма информации первой страницы и информации второй страницы; информация всей книги — это сумма информации всех ее страниц.

Впрочем, здесь мы переходим уже не в область математики, а в область другой научной дисциплины — математической лингвистики.

После того, как Клод Шеннон заложил основы вероятностной теории информации, эта теория нашла отклик среди ученых различных специальностей: биологов, лингвистов, философов, генетиков, искусствоведов, психологов, экономистов, геологов, химиков, математиков. Кодом стали называть любую систему знаков, предназначенных для передачи сообщений. Термины теории информации получили широчайшее применение в самых разных публикациях.

Но вот выходит краткая статья самого создателя этой теории, Клода Шеннона, озаглавленная «Бандвагон». Этим словом в США называют политическую партию, добившуюся популярности и победившую на выборах, или просто группу лиц, программа которых находит у населения широкую поддержку. Родился этот термин, вероятно, потому, что обычно победивший на выборах кандидат проезжал по городу в открытой машине, сопровождаемый оркестром (английское band значит оркестр, джаз, а wagon — повозка, карета).

За последние годы теория информации превратилась в своего рода бандвагон от науки — так начинает свою статью Шеннон. Появившись на свет в качестве специального метода в теории связи, она заняла выдающееся место как в популярной, так и в научной литературе. А в результате «значение теории информации было, возможно, преувеличено и раздуто до пределов, превышающих ее реальные достижения».

Очень редко удается открыть одновременно несколько тайн природы одним и тем же ключом, предостерегает Шеннон. Здание нашего искусственно созданного благополучия слишком легко может рухнуть, если в один прекрасный день окажется, что при помощи нескольких магических слов вроде информация, энтропия, избыточность и т. п. нельзя решить всех нерешенных проблем.

«Что можно сделать, чтобы внести в сложившуюся ситуацию ноту умеренности?» — задается вопросом сам Шеннон. И отвечает так: прежде всего представителям различных наук нужно ясно понимать, что основные положения теории информации касаются очень специфического направления, что оно не обязательно должно оказаться плодотворным, скажем, в психологии или

экономике. «Я лично полагаю, что многие положения теории информации могут оказаться очень полезными в других областях науки,— говорит Шеннон.— Действительно, уже достигнуты некоторые весьма значительные результаты. Однако поиск путей применения теории информации в других областях не сводится к тривиальному переносу терминов из одной области науки в другую. Этот поиск осуществляется в длительном процессе выдвижения новых гипотез и их экспериментальной проверке».

Эти слова основоположника теории информации прекрасно иллюстрируют ситуацию, что сложилась в языкознании, когда человеческую речь стали рассматривать в качестве своеобразного кода. Изучение языка как кода началось уже на заре теории информации. Но только сейчас, по мере того как растут наши знания, мы начинаем постигать, насколько сложен, специфичен, можно сказать, удивителен код — наш язык, с помощью которого мы общаемся.

### БИТЫ И БУКВЫ

Сколько информации несет одна буква? Вопрос этот возник в первые же годы рождения теории информации. Простейший ответ на него найти легко. Надо взять число букв в том или ином алфавите мира, а затем выразить его в двоичных логарифмах. Ведь каждая буква — это один из возможных исходов, вроде выпадения герба или решетки монеты, одной из граней кубика и т. д. Число же этих исходов равно числу различных букв.

В русском алфавите тридцать три буквы, плюс еще нулевая — пробел между словами. Но так как *e* и *ё* сейчас практически не различаются, то их можно считать одной буквой. Можно объединить мягкий знак с твердым. В итоге получаем тридцать две буквы, тридцать два кодовых знака нашего языка. Двоичный логарифм тридцати двух равен пяти. Получаем величину в пять бит. Столько информации несет одна буква нашего алфавита.

В языках, пользующихся латинским алфавитом, число букв равно двадцати шести. Прибавим еще пробел, нулевую букву, а затем выразим это в двоичных логарифмах.

рифмах, то есть в битах. Получаем величину в 4,76 бит. Столько информации несет одна буква английского, испанского, немецкого, французского языков.

В алфавите самоанского языка, на котором говорят жители одного из полинезийских архипелагов, всего лишь шестнадцать букв. Двоичный логарифм шестнадцати равен четырем, значит, одна буква самоанского алфавита несет информацию в четыре бита, на один бит меньше, чем буква русского.

Все это было бы совершенно правильно, если бы буквы алфавита встречались с одинаковой вероятностью, если бы языки не имели свойства, называемого в теории информации избыточностью. Избыточность позволяет определять, насколько разнится максимальная информация, которую может нести знак кода, от реальной, которую он несет. Иными словами, это своеобразная мера «неэкономности» кода, в том числе и любого языка мира.

Интуитивно, не производя подсчетов, мы догадываемся, что одни буквы, вроде *о* или *с*, встречаются часто, другие, вроде *ф* или *э* — реже. То же самое и со словами: союз *и* или местоимение *я* весьма часты, а прилагательное *сиятельный* или глагол *выковыривать* — гораздо реже. Мы согласуем нашу речь в роде, числе, падеже, залогах, наклонениях и т. д. Наша речь осмысленна, значит, не всякое слово может соединяться с другим... Короче говоря, на язык наложены ограничения, начиная с употребительности отдельных букв и кончая сочетаемостью слов по смыслу.

Подсчитать частоту употребления букв в различных алфавитах мира нетрудно. Выше приводилась величина информации, которую несет одна буква различных языков мира. Но если учитывать, что буквы имеют разную вероятность появления в тексте, то эта величина будет иной. Причем даже в языках, которые пользуются одним и тем же латинским алфавитом. В немецком она равна 4,1 бита, в английском — 4,03 бита, в испанском — 3,98 бита, во французском — 3,96 бита (исходная величина для всех них была, как вы помните, 4,76 бита). Для русского языка величина информации, которую несет одна буква, сократилась до 4,35 бита, для самоанского — до 3,4 бита.

Но это — лишь первое приближение. Ведь различные вероятности имеют не только отдельные буквы, но

и сочетания двух, трех и т. д. букв (в русском языке, например, невозможно сочетание гласная плюс мягкий знак и другие подобные сочетания). Если учесть вероятности двухбуквенных сочетаний, то величина информации, которую несет одна буква русского алфавита, будет равна 3,52 бита, английского — 3,32 бита, французского — 3,17 бита, а самоанского — 2,68 бита. Учет вероятностей трехбуквенных сочетаний для русского алфавита дает величину в 3,01 бита, для английского — 3,1 бита, для французского — 2,83 бита.

Вы, вероятно, заметили, что по мере того, как мы учитываем все более и более «дальние» связи между буквами, уменьшается количество информации на одну букву и соответственно увеличивается избыточность языка. Возможно, вы заметили и другое: одновременно происходит сближение величины информации в языках, пользующихся различными алфавитами, по мере того, как мы все точнее и точнее описываем наш код-язык.

Действительно, когда была определена величина информации, приходящаяся на одну букву осмысленного текста в разных языках мира, она оказалась примерно одинаковой.

### МЕТОДОМ ОТГАДЫВАНИЯ

Каким образом удалось определить эту величину? Ведь учет даже трехбуквенных сочетаний, частоты их употребления в текстах того или иного языка требует работы огромного объема. Между тем статистические связи между буквами, конечно, не ограничиваются триграммами, учетом частот всех этих *тс*, *ств*, *щий* и т. п. Надо знать частоты сочетаний четырех, пяти и более букв.

Однако на этом пути мы заходим в тупик. Число возможных комбинаций из девяти букв, например, измеряется триллионами. Сколько же текстов мы должны исследовать, чтобы получить достоверную статистику! А ведь и девять букв — не предел. Мы пишем и говорим предложениями, значит, в пределах этого предложения будут существовать статистические связи между буквами, задаваемые лексикой, грамматикой, синтаксисом. Длина же предложения измеряется десятками букв.

Задача, казалось бы, неразрешимая... Однако Клод Шеннон, одним из первых занявшийся изучением языка

как кода, предложил остроумный способ решения этой задачи с помощью так называемого метода отгадывания. Метод этот был усовершенствован крупнейшим математиком нашей страны А. Н. Колмогоровым. С помощью этого метода была определена величина информации, приходящаяся на одну букву русского, английского, французского, польского, румынского, азербайджанского и других языков мира. (Первоначально опыты проводились в МГУ, а затем были продолжены ленинградским лингвистом Р. Г. Пиотровским, описавшим их в книге «Информационные измерения языка».)

Мы уже говорили о том, что информация — это то, что снимает неопределенность, уничтожает незнание... Где существует неопределенность, которую снимают буквы того или иного языка? Разумеется, в голове читателя осмысленного текста на этом языке. Вот этого-то читателя и надо использовать в качестве своеобразного измерителя величины информации, которую несет одна буква алфавита.

На одном полюсе будет человек, который не знает языка. Для него все буквы равновероятны, избыточность алфавита равна нулю, и один знак несет максимальную информацию кода (и никакой осмысленной информации не дает!). На другом полюсе — носитель языка, превосходно знающий его, интуитивно чувствующий вероятность появления того или иного слова в тексте. Такому человеку дается определенный текст, он прочитывает несколько страниц, вникает в стиль, сюжет, манеру повествования. Затем текст закрывается, и далее начинается процедура отгадывания. Предлагается найти продолжение этого текста: какая буква должна появиться?

Например, в опытах Пиотровского предлагался отрывок из статьи, опубликованной в газете «Известия»: «И вот, приехав в тот город, где друг мой в свое время справлял свадьбу, я встретил его жену и сына, который учится уже в пятом классе. И сын его меня спрашивает, правда ли, что я служил с его отцом, на каких кораблях плавали, где бывали. Я отвечаю на эти как будто нехитрые на первый взгляд...»

Дальше текст закрывался и предлагалось угадать букву, которая должна последовать. Отгадчик мог отказать от ответа (в самом деле, мало ли какое слово

может последовать после слов «на первый взгляд») или мог предложить какую-либо букву (например, *в*, считая, что следующим словом будет *вопросы*). Вслед за тем отгадчику называлась эта буква (в данном случае ею была буква *р*). Процедура вновь повторялась, причем отгадчик мог называть предполагаемую букву с малой или большой степенью уверенности, назвать две возможные в данном контексте буквы, имеющие, по его мнению, равную вероятность, или же отказаться от угадывания (в нашем примере требовалось угадать слово *ребячьи* — угадывание его шло безошибочно, когда известны стали три первые буквы; вероятно, столь же нетрудно угадать и дальнейшее продолжение — слово *вопросы*).

Полученные результаты обрабатывались математически. В итоге можно было получить величину информации, приходящуюся на одну букву осмысленного русского текста. Опыты, проведенные в МГУ, показали, что для классической русской прозы (отгадывались «Детские годы Багрова-внука» Аксакова и «Литературные вечера» Гончарова) она равна примерно одному биту. Сходную величину дали и опыты Р. Г. Пиотровского. Если сравнить эту величину в пятью «исходными» битами, мы увидим, что избыточность литературного русского языка классической прозы равна восьмидесяти процентам. Иными словами, четыре из пяти букв — «лишние»!

В пределах семидесяти-восьмидесяти процентов оказалась величина избыточности и для английского, французского, румынского, польского, азербайджанского, казахского языков. Весьма вероятно, что и во всех других языках мира избыточность, которую называют еще своеобразным запасом прочности языка, находится в этих пределах. Если бы в языке не было избыточности, то любое сочетание кодовых знаков, букв, образовало бы осмысленное слово, и тогда почти невозможно было бы исправить ошибку в написании. Не существовало бы и правил грамматики, и сочетаний смыслов слов... Так что избыточность это не излишество, не показатель несовершенства языка по сравнению с техническими кодами, а его важное и нужное свойство, которое возникло многие тысячи лет назад, но лишь сейчас начинает нами осознаваться благодаря теории информации.

Во всяком ли русском тексте четыре пятых букв — «лишние»? Нет, не во всяком. Вспомните телеграммы: в текстах телеграмм мы стараемся уменьшить избыточность языка, опускаем предлоги, глагольные связки и т. п. Однако эта экономия имеет свой предел, иначе получатель нашей телеграммы не поймет текст. Ленинградские исследователи О. Л. Смириов и А. В. Екимова, используя метод отгадывания, попробовали определить величину информации, приходящуюся на одну букву «телеграфного языка» (или если применять лингвистическую терминологию, подъязыка телеграмм). Для этого были взяты телеграфные тексты общим объемом в пятнадцать тысяч букв. Оказалось, что величина информации почти в полтора раза превышает ту, что несет одна буква литературного русского языка!

На другом полюсе находятся деловые тексты. Величина избыточности здесь, в отличие от телеграмм, наоборот, завышается для большей надежности. Деловые тексты имеют стандартные обороты, словосочетания, синтаксические конструкции. Да и содержание их также вращается в строго заданной сфере — будь это химия, геология, метеорологическая сводка или бухгалтерский отчет. Так, например, опыты по угадыванию специального текста — документации финансового отдела одного из рудоуправлений Донецкой области — показали, что величина избыточности здесь доходит до девяноста процентов.

Правда, девять из десяти букв текста документации угадывали только специалисты, хорошо знакомые с бухгалтерским делом, имевшие опыт работы с подобного рода текстами изо дня в день. Когда же в эксперименте приняли участие специалисты, знакомые с текстом в меньшей степени, они смогли угадать не девять, а восемь букв из каждых десяти (то есть избыточность равна не девяноста, а восьмидесяти процентам). А неспециалисты могли угадать лишь четыре — шесть букв из каждых десяти букв этого бухгалтерского текста, в котором встречалось много обособленных слов, сокращений, стандартных формулировок, понятных лишь знакам.

Здесь мы переходим к интереснейшему вопросу, связанному уже не только с языком-кодом, но и тем, как

пользуется получаемой информацией человек. Один и тот же текст может содержать разное количество информации в зависимости от того, кто получает эту информацию: насколько владеет языком получатель, каков его прошлый опыт, какова его психологическая настройка и т. п. Иными словами, от информации объективной мы переходим к информации субъективной.

Работы в этом направлении только-только начинаются. С помощью теории информации делаются попытки выразить в числах различные аспекты деятельности человека, которые казались прежде чисто субъективными, неподвластными мере и числу. Например, восприятие, понимание, обучение, запоминание, принятие решений и многие другие психические процессы.

Причем делается это уже не в стиле бандвагона от науки, о котором иронично писал Шеннон, а на основании детальных и серьезных исследований. На первых этапах терминами, взятыми из арсенала теории информации, назывались явления, порой давным-давно известные специалистам в данной области, лишь получавшие модное наименование. Ныне речь идет о точных мерах, не об «информации вообще», которую перерабатывает мозг, а о величине, выраженной в битах, определенной по формуле Шеннона.

Идеи и методы теории информации оказались плодотворными не только в языкознании и психологии, но и в биологии. Они позволили биохимикам расшифровать сложнейший код, посредством которого передается наследственная информация. Находят применение эти методы и в изучении искусства, прежде всего поэтического. Ведь материалом для поэта служит язык со всеми его статистическими закономерностями.

### ФОРМУЛЫ СТИХА

Первые исследования русского стиха методами математики были проведены в начале нашего века. В 1910 году известный поэт и теоретик Андрей Белый опубликовал свои работы, посвященные статистическому анализу четырехстопного ямба. В двадцатые годы математику в стиховедении успешно применяли профессиональные литературоведы Г. Шенгели, В. Чудовский, Б. Ярхо и особенно Б. Томашевский.

Новый этап начался в 1960 году, когда идеи теории

информации и кибернетики начали применяться в изучении стиха. Во главе этого направления стал А. Н. Колмогоров. В настоящее время над математическим анализом стиха работают как профессиональные стиховеды и лингвисты, так и математики. Используя богатый опыт русских стиховедов, «стихOMETрам» удалось внести в поэтику математические характеристики, на основании которых можно точно судить о богатстве и своеобразии ритмики различных русских и советских поэтов. Не так давно вышла в свет монография М. Л. Гаспарова «Современный русский стих», где на основании статистических подсчетов дан подробный анализ таких современных размеров, как дольник, тактовик, акцентный стих, основоположником которого считается Маяковский.

Делаются попытки применить математику в изучении более «высокого этажа», чем ритмика, — рифмы и звуковой инструментовки стиха. И, что самое интересное, начинают ставиться вопросы, в традиционном литературоведении даже не возникавшие. Например, об информационной «емкости» стиха (тех, кто интересуется «стихOMETрией» отошлем к книге Гаспарова, вышедшей в 1974 году в издательстве «Наука», а также к статьям, публиковавшимся в журнале «Вопросы языкознания», и брошюре автора этой книги «Математика и поэзия», выпущенной издательством «Знание» в 1962 году).

Речь подчиняется определенным статистическим закономерностям. Поэзия накладывает на язык еще ряд ограничений: требования ритмики, рифмовки, звуковой инструментовки и т. д. Подсчеты показывают, что современный стих более «легок», чем классический, в отношении ритмических требований. Вероятность случайного возникновения в обычной речи строки четырехстопного ямба равна примерно трем сотым — в последовательности из ста русских слов автоматически может образоваться в среднем три строки четырехстопного ямба. Вероятность случайного образования современного четырехдольника (которым написаны «У самого моря» Ахматовой, «Люблю» Маяковского, «Февраль» Багрицкого, «Победитель» Симонова, пьесы В. Гусева и многие другие произведения) равна двум десятым — уже не три, а двадцать строк может автоматически образоваться в последовательности из ста русских слов. Вероятность же образования «акцентного», четырехудар-

ного стиха равна единице, ибо любая последовательность из четырех русских слов может служить строкой этого стиха, безразличного к числу безударных слогов в промежутке между ударными.

Но раскрепостив стих по ритму, сделав его «легче» для языка, современное стихосложение налагает более строгие ограничения по рифме. «Грамматические» рифмы (типа *стоять — лежать, окном — пером* и т. п.) образовать легко. А рифмы этого типа практически не встречаются в современном стихе. Его организует рифма, это по преимуществу рифменный стих — вспомните слова Маяковского о том, какую большую роль в его творчестве играют поиски глубоких и неожиданных рифм. Сняв одни ограничения — ритмические, поэты налагают другие ограничения — в области рифмовки. И все эти ограничения можно выразить на языке точных чисел.

Сравнение «степени трудности» различных стихотворных размеров логичнее всего проводить, вычисляя информационные «траты» на ритм и рифму, а не просто сопоставляя вероятности их случайного возникновения в прозе (так, в упомянутой выше брошюре «Математика и поэзия» автор ошибочно полагал, что требования классической рифмовки от десяти до ста раз «труднее», чем требования рифмовки современной, а требования ритма соответственно от десяти до ста раз «легче»).

Затраты информации на образование двух строк четырехстопного ямба равны примерно десяти битам. Затраты на связывание этих ямбических строк классической рифмой равны семи битам.

Обратимся к современному стиху. Затрата на ритм в четырехдольнике равна пяти битам, то есть в два раза меньше, чем в ямбе. А на рифму? Так как «грамматические» рифмы современными поэтами не употребляются, траты будут большими, чем в классическом стихе — не семь, а восемь бит. Суммируем траты на ритм и рифмы: в классическом четырехстопном ямбе это составит семнадцать бит, в современном четырехдольнике — тринадцать бит.

Вполне понятно, что у каждой поэтической школы, большого поэта и, быть может, отдельного произведения имеется свое специфическое распределение «трат» информации на ритм, рифму, звуковую выразительность. Однако, несмотря на все эти траты, в распоряжении поэта есть достаточно средств, чтобы выразить

свои мысли, чувства, свою «модель мира»... «Поэт может вложить в сообщение из 400 букв (сообщение чисто «цифровой природы», несущее информацию порядка тысячи битов, т. е. количественно ничтожную с точки зрения современной техники), — пишет академик Колмогоров, — целый мир чувств, который справедливо признается не поддающимся формализации в понятиях, и создать с такими скромными средствами «канал связи» непосредственного общения со своими современниками и потомками, раскрывающий, разрывая ограничения пространства и времени, его неповторимую индивидуальность».

Мы рассказывали о том, как была определена величина информации, которую несет одна буква русского текста. Она равна примерно одному биту, причем меняется в зависимости от того или иного стиля.

Опыты по определению количества информации, содержащейся в одной букве поэтического текста, показали, что величина эта превосходит величину, которую несет одна буква прозаического текста.

На первый взгляд это кажется странным. Ведь речь поэта подчиняется правилам ритма, она рифмована и т. п. (и выше мы приводили величины «трат» по этим правилам). Однако поэтический синтаксис более свободен, чем синтаксис прозы. Да и сочетания слов в поэзии более свободны, неожиданны. Кроме того, наша речь позволяет нам выражать мысли многими способами. В обычной практике мы безразличны к этому множеству. Однако в художественной речи люди сознательно используют богатые возможности, которые предоставляет язык. Это означает, что при сохранении главного требования речи — передачи смысла — достигается еще и дополнительное воздействие через ритм, звучание рифм, звуковую инструментовку. То, что в обычном языке «пропадает зря» (например, ритмическое построение фразы, скопление однородных звуков, случайная рифмовка, возникающая порой в нашей прозаической речи), в поэзии находит применение, и благодаря этому художественная речь несет больше информации, чем речь нехудожественная.

Разумеется, это относится к стихам больших поэтов, умело и творчески использующих возможности, которые предоставляет им родной язык. Когда же мы имеем дело с ремесленными поделками, то здесь величина

информации, которую несет одна буква стиха, гораздо меньше той, что несет одна буква прозы. На кафедре теории вероятностей МГУ сопоставили информационную емкость «Поединка» Куприна и весьма скромного по своим поэтическим высотам стихотворения, напечатанного на обороте листка отрывного календаря. Оказалось, что проза в два раза более емка, чем отштампованный по шаблону стих!

### СКОРОСТЬ УСТНОЙ РЕЧИ

Поскольку основное средство общения — разговорная речь, то, вероятно, важно знать не только число бит, приходящихся на тот или иной звук, но и скорость передачи информации при разговоре.

И в нашей стране, и за рубежом было проведено множество исследований, посвященных этой теме. Кодовой единицей для письменного текста является буква. Для устной речи — фонема, своеобразный атом звукового языка (о нем рассказывает очерк «Формулы фонемы»). Попыты по угадыванию фонем, подсчеты частоты их употребления, наконец, анализ спектрограмм фонем и сравнение их с изображениями букв в машинописном тексте — все это позволило определить величину избыточности звукового кода языка. Она оказалась примерно такой же, как и величина избыточности алфавитного кода. Только в зависимости от «подъязыка», от стиля и характера разговора эта избыточность может варьировать в еще больших пределах, чем «запас прочности» письменного текста.

Возьмем разговор двух друзей, понимающих, как говорится, друг друга с полуслова. В их речи будут умолчания, намеки, пропуски слов и «съедания» отдельных звуков и даже грамматических окончаний (в письменной речи это недопустимо). Вполне понятно, что избыточность такого разговора будет, пожалуй, еще меньше, чем избыточность телеграфного стиля, о котором мы рассказывали.

А вот другой разговор: диалог между дежурным на аэродроме и пилотом, находящимся в воздухе. Во-первых, этот разговор по радио имеет определенные стандарты. Во-вторых, тематика его ограничена. В-третьих, он происходит в условиях шума, который создает самолет, поэтому приходится повторять нерасслышанные или

нечетко услышанные слова и фразы. В-четвертых — и это самое главное,— ошибка в одно слово может стоить жизни пилоту. В итоге избыточность такого разговора равна девяноста шести процентам. Иными словами, из ста слов пилот и диспетчер предпочитают говорить девяносто шесть «лишних», зато это гарантирует им надежность связи и, стало быть, безопасность жизни пилота, ведущего самолет.

Скорость передачи информации при таком разговоре очень невелика. Она равна примерно двум десятым бита в секунду. Обычная скорость передачи информации при разговоре примерно в двадцать — тридцать раз больше, она равна пяти — шести битам. Но это при нормальном темпе речи. Как известно, этот темп можно значительно увеличить: очень медленная речь почти в пять раз медленней, чем очень быстрая. Сколько же информации за секунду может передать человек при очень быстром разговоре?

Вот любопытный расчет. Человеческое ухо может воспринять, а человеческий рот произнести чудовищно много различных звуков. Пропускная способность человеческого уха, то есть количество информации, которое может быть передано в единицу времени, равна пятидесяти тысячам бит в секунду (по телефону мы можем передать только тридцать тысяч бит в секунду, он сужает диапазон). Величина эта огромна. Но, конечно, мозг получает не всю звуковую информацию, он ее фильтрует, просеивает сквозь сито фонем.

Сколько же информации доходит до мозга? Насколько частым является это сито? Оказывается, оно очень частое. Настолько частое, что количество информации уменьшается в тысячу раз: уже не пятьдесят тысяч, а просто пятьдесят бит в секунду получает мозг, воспринимая предельно быструю речь.

Но и это величина немалая. Опыты показывают, что большее количество информации наш мозг и не в силах обработать сознательно (например, опытная машинистка или пианист передают сигналы со скоростью двадцать пять бит в секунду). Таким образом, скорость передачи информации при предельно быстром разговоре (сорок — пятьдесят бит в секунду) в двести — триста раз превышает скорость передачи информации при разговоре диспетчера с летчиком и в десять раз скорость разговора в нормальном темпе.

## ИЗМЕРЕНИЕ ФОНАЦИИ

Главная задача письменной речи — передача смысла. В устной речи, кроме смысловой, содержится еще немало «несмысловой» информации: в интонации, громкости, индивидуальных особенностях голоса говорящего и т. д. Иногда же она может даже противоречить информации смысловой. Например, когда мы говорим иронически «Спасибо!» или «Здравствуйте!» Очень часто мы больше доверяем интонации, с которой произносится фраза, чем прямому смыслу слов этой фразы.

Изучение несмысловой информации важно для техники связи. Ведь при трансляции театральных постановок, музыкальных концертов и т. д. нам нужно передавать как раз эту несмысловую информацию. Да и при передаче чисто деловых текстов по радио интонация, тембр голоса, громкость речи также играют большую роль: вспомним голос Юрия Левитана, когда он читает важное правительственное сообщение.

Можно ли оценить несмысловую информацию, что содержится в речи? Более двадцати лет назад немецкий специалист по технике связи К. Кюпфмюллер попытался сделать это. Он разбил несмысловую информацию, содержащуюся в устной речи, на три большие группы. Первая — информация, что содержится в интонации, вторая — информация, связанная с индивидуальными особенностями голоса, третья — информация, доставляемая громкостью речи. Сумма этих трех величин давала несмысловую информацию, которая сравнивалась затем с содержащейся в той же речи смысловой информацией. И вот какие результаты были получены при таком сравнении.

При нормальном разговоре дополнительная, то есть несмысловая информация, что содержится в интонации, громкости, индивидуальных особенностях голоса, не превышает обычно семидесяти пяти процентов от информации смысловой. При очень быстром разговоре она составляет не более чем тридцать процентов от смысловой информации. А при очень медленной беседе — не более ста пятидесяти процентов.

Чем объясняется такая разница в цифрах? Почему при медленном разговоре несмысловая информация может в полтора раза превышать смысловую, а при быстром — быть в три раза меньше?

Видимо, потому, что при быстрой речи гораздо труднее следить за всеми оттенками речи. Да и узнать голос человека при этом гораздо труднее. Ведь наш мозг не может одновременно воспринимать всю смысловую и несмысловую информацию, которую несет человеческий голос при быстром разговоре. Главная задача нашей речи — передача смысла. Когда речь становится слишком быстрой, мы поневоле отказываемся воспринимать несмысловую информацию и направляем сознание только на восприятие смысла. А при медленном разговоре мы спокойно можем воспринять все оттенки интонации, индивидуальные особенности голоса и различные его градации по громкости.

Разумеется, работа Кюпфмюллера была лишь самым первым приближением к тем поистине неисчерпаемым возможностям, которые дает изучение человеческой речи методами теории информации. В последние годы оформилась новая область исследования. Ее предмет — неязыковые средства общения. Называют эту область исследования паралингвистикой. И пальма первенства принадлежит тут явлению, которое принято называть фонацией.

Говорят, что древнегреческий философ Сократ высказывал свое мнение о человеке лишь тогда, когда слышал его голос. Арабский ученый Абуль-Фарадж, живший в XIII столетии, полагал, что тот, «кто разговаривает, постепенно снижая голос, — несомненно чем-то глубоко опечален... кто говорит слабым голосом — робок, как ягненок, тот, кто говорит пронзительно и несвязно, — глуп, как коза». А вот что говорит В. П. Морозов в своей книге «Тайны вокальной речи»: «Человек произносит слова. Мы воспринимаем их смысл. Но как много смысла, помимо слов, кроется в самом звуке голоса! Прислушайтесь к звукам речи незнакомого человека... Разве тембр голоса, манера говорить, интонации не расскажут вам многое о его чувствах и характере? Ведь голос бывает теплый и мягкий, грубый и мрачный, испуганный и робкий, ликующий и уверенный, ехидный и вкрадчивый, твердый, живой, торжествующий и еще с тысячью оттенков, выражающих самые разнообразные чувства, настроения человека и даже его мысли».

Необычайно сложно измерить в битах все эти тысячи оттенков человеческого голоса. А ведь они, вне всякого сомнения, несут информацию и весьма существенную

при разговоре. Как-то Британское радио провело эксперимент: в эфире выступило девять чтецов, а слушателям было предложено определить не только возраст и пол этих чтецов, но и профессию. Безошибочно были распознаны актер, судья и священник — голос их имеет профессиональную постановку, причем она различается у актера, священника и судьи.

Голос человека обладает тембром, который называют образно колоритом, окраской звука и даже его цветом. По тембру мы легко можем отличить знакомый голос или тип голоса певца. Это — индивидуальное свойство голоса, оно не связано прямо с системой языка (хотя, вне всякого сомнения, тембр речи несет информацию, иначе мы бы не узнавали голоса по тембру). Однако есть в звучащей речи признаки, которые характеризуют язык как систему. Это так называемые тона. Слова в таких языках, называемых тоновыми, различаются не только звуками, но и тоном, с каким они произносятся.

Тон может быть высоким или низким, восходящим или нисходящим. В некоторых языках, например вьетнамском, число разных тонов доходит до шести. Тоновые языки распространены в Юго-Восточной Азии, в Западной Африке, системы тонов имеет китайский язык, на котором существует древнейшая письменность, и бесписьменные наречия бушменов. В тех же языках, где системы тонов нет, существует интонация. И она, как правило, тоже связана не только с речью конкретного человека, но и системой языка.

Например, во многих языках, в том числе русском, большую роль играет логическое ударение. С его помощью мы выделяем слово, которое хотим подчеркнуть или сместить смысловой акцент, переместить члены предложения, как бы перегруппировать смысл фразы.

Возьмем простейшую фразу: «Я еду в Москву». У нее может быть разный смысл, в зависимости от того, на какое слово мы поставим логическое ударение; именно я (а не ты) еду в Москву; я еду в Москву, а не лечу самолетом; я еду в Москву, хотя собирался в Киев...

Английский связист Дж. Берри сделал попытку вычислить количество информации, которое несут такие логические ударения в английском языке. Он проанализировал магнитофонные записи типичных телефонных

разговоров. Выяснилось, что ударение чаще всего падает на наиболее редкие слова. Применяв формулу Шеннона и взяв данные частотных словарей, французский ученый Б. Мандельброт вычислил, что средняя величина информации, которую мы получаем от логического ударения, близка 0,65 бита на слово.

Но ведь это — лишь один из компонентов, связанных с той информацией, которую несет интонация нашей речи. На VII Международном конгрессе по фонетике, состоявшемся в Монреале в 1971 году, французский лингвист Пьер Леон выступил с докладом «Где изучают интонацию?» Ученые выделяют две основные функции интонации — лингвистическую, о которой речь шла выше, и нелингвистическую. Но, говорит Леон, «один и тот же вид информации дается всеми системами, присутствующими одновременно, и любая из них может преобладать над другой».

С помощью интонации человек может выражать свое внутреннее состояние сознательно или непроизвольно. Интонация позволяет чтецам и актерам передавать тончайшие оттенки смысла слов и отдельных фраз. Интонация играет большую роль в грамматическом оформлении речи (логическое ударение, завершенность, фразовое ударение и т. д.). Интонация характеризует определенные стили речи и языка: недаром мы говорим об ораторской, снобистской, фамильярной, изысканной, казенной, дружеской интонации. Интонация выражает наши эмоции — радость, страх, грусть, гнев...

Пьер Леон экспериментально установил, что для выражения грусти служат следующие признаки, выраженные в терминах акустики: узкий мелодический интервал; ровный мелодический контур; медленный темп; слабая интенсивность. Для выражения гнева интенсивность будет не слабой, а сильной и т. д. Словом, ученые начинают с помощью современной акустической техники выявлять параметры, по которым передается интонационная информация.

Интуитивно же мы все прекрасно воспринимаем эту информацию. В одном из экспериментов фразу, произнесенную с определенной интонацией, пятьдесят восемь человек из ста опознали как выражение сомнения, а сорок восемь — как иронию. Когда же эта фраза была произнесена со вздохом, ее единогласно признали выражением грусти.

Мы могли бы очень долго рассказывать об интонации, о чудесах, которые творят с ее помощью актеры и чтецы, об удивительном мире звучащей речи. Но об этом достаточно много написано и в популярной литературе, и в научных трудах, и в мемуарах (прочитайте, например, «Мою жизнь в искусстве» Станиславского, и вы найдете там десятки самых ярких примеров артистической фонации). Наша цель была другой: показать, насколько сложно и трудно измерить информацию, которую несет наша разговорная речь, насколько еще грубы и прикидочны оценки этой информации в битах.

А ведь это еще далеко не все сложности, которые возникают при описании языка как кода. «В настоящее время едва ли можно мыслить себе лингвистическое исследование без учета двух противоположных процедур: устранения избыточности и использования избыточности», — так говорил профессор Р. О. Якобсон, подводя итоги IX Международного конгресса лингвистов.

При разговоре мы ориентируемся на контекст, на собеседника, на ситуацию, мы пользуемся не только интонацией, но и жестами, телодвижениями, мимикой, сопровождающими нашу речь. «Дай мне!» — говорим мы, указывая пальцем на предмет, не называя его. «Это?» — спрашивает собеседник, протягивая соседний предмет. «Нет, это», — отвечаем мы, указывая на нужный. В разговоре, казалось бы, опущено самое важное — наименование предмета. Тем не менее мы понимаем друг друга и передаем нужную информацию.

В экстренных случаях мы сокращаем нашу речь до предела, ограничиваемся одним-единственным словом вместо фразы, сказав, однако, все, что требовалось в данной ситуации. Вспомните восклицания вроде «Тревога!», «Пожар!», «Вор!», «Сюда» и т. п. В любом другом коде это было бы невозможно.

Для описания грамматики любого человеческого языка требуется объемистый том, а то и не один. И, что опять-таки делает язык уникальным кодом, нарушение правил грамматики все-таки позволяет нам понимать друг друга. Иностранец, плохо владеющий русским языком, спросит у нас: «Троллибас берет Астория, пожалуйста?» — и мы поймем, что гостю Ленинграда надо проехать к гостинице «Астория». Попробуйте-ка нару-

шить правила сочетаний знаков любого другого кода, и вы получите либо бессмыслицу, либо совсем не то, что хотели бы выразить.

Не менее поразительное свойство языка как кода — это возможность свободно сочетать значения слов. «Все человеческие языки — китайский или аранта, современный английский или неизвестный язык кроманьонца — являются по определению семантическими кодами — условными моделями, зрительными или словесными, представляющими заранее согласованные между членами коллектива значения», — так американские ученые Дж. Перри и А. Кент, создатели «семантического кода» для электронных вычислительных машин, характеризуют язык человека.

Но эта характеристика не совсем верна. Во-первых, никто никогда не договаривался — ни китаец с австралийцем аранта, ни англичанин с кроманьонцем, — что одно слово будет значить то-то, а другое то-то. Да и на каком, собственно говоря, языке они должны были договариваться о значениях слов? Во-вторых, в отличие от семантических кодов, предназначенных для ЭВМ, сочетания смыслов в нашем языке не подчиняются строго определенным правилам. Иначе мы не могли бы сказать ничего принципиально нового ни в жизни, ни в науке, ни в искусстве.

Вспоминается такой курьез. На одной из конференций по лингвистике приводились образцы фраз, правильных грамматически, но не имеющих смысла. Однако для каждой из этих фраз удавалось найти контекст, в котором она становилась осмысленной! Даже для классической фразы «Идея яростно спит», попавшей во все работы по современной лингвистике как образец бессмысленной, но грамматически правильной.

Доктор филологических наук И. И. Ревзин предложил сделать фразу первой строкой четверостишия:

Идея яростно спит,  
Ворочается во сне...  
Идея в висках стучит,  
Нашептывая мне.

И, казалось бы, бессмысленная фраза сразу же становится осмысленной в контексте четверостишия. Шутки ради добавим, что фразу эту можно осмыслить и не только с помощью поэзии. Если «Идею» считать

женским именем (а такие имена давались многим девочкам в двадцатых — тридцатых годах), то «Идея яростно спит» будет означать, что девушка по имени Идея спит, тяжело и глубоко дыша, разметавшись на подушках, — словом, «яростно».

Таким образом, ни в грамматике, ни в семантике языка нет таких строгих ограничений, как в остальных кодах, которыми пользуется ученый, инженер, телеграфист и т. д. Нет правил без исключений — таков наш язык. Между тем в других кодах правила строги и однозначны, исключений они не признают — па то они и коды!

В языке, в отличие от кодов, один и тот же смысл может быть выражен различными словами или фразами.

Одно и то же слово может иметь несколько смыслов. Слово может иметь антоним, своего антипода по смыслу. Все это обогащает язык, делает его гибким, живым, выразительным. Кстати сказать, в одном из фантастических романов описывается диктатура, типа фашистской, лидеры которой вводят новый язык. На языке этом нельзя сказать ничего такого, что противоречило бы догмам. Синонимы и даже антонимы устраняются: к чему иметь слово плохо, когда можно сказать нехорошо? Устраняются и прилагательные типа прекрасно, восхитительно, великолепно — их заменяют обороты очень хорошо и очень-очень хорошо... Человеческий язык переделывается в технический код для того, чтобы превратить людей в послушных диктатуре рабов-роботов.

Почти любое слово имеет несколько значений. И, как считает академик Л. В. Щерба, «в нашем повседневном употреблении мы скатываемся на формальную точку зрения, придавая слову *слово* значение «фонетического слова»... Это, в сущности говоря... даже просто типографская точка зрения. На самом деле мы имеем всегда столько слов, сколько данное фактическое слово имеет значений (так и печаталось, между прочим, в старых словарях: заглавное слово повторялось столько раз, сколько у него было значений)».

Словарь омонимов русского языка содержит несколько тысяч слов. Загляните, однако, в любой толковый словарь нашего языка (да и любого другого), и вы увидите, что каждая словарная статья дает, как правило, не

одно, а два и более значений слова. А ведь омонимия, совпадение знаков по форме, имеющих разное значение, исключается в любом коде — любом, кроме нашего разговорного языка!

Мы всегда можем в случае необходимости сокращать избыточность. В устной речи, особенно когда мы говорим торопливо, опускаются не только связки и служебные слова, но и многие звуки внутри, в начале или в конце слова. Сан Саныч вместо Александр Александрович, чезк вместо человек...

В письменной речи мы пользуемся сокращениями, так называемыми аббревиатурами: НТО, НТР, ВЦСПС, РСФСР и много, много других (существуют даже специальные словари аббревиатур для некоторых языков мира).

Для технических кодов, разумеется, такие аббревиатуры невозможны хотя бы потому, что сочетания букв, их образующие, с точки зрения обычного языка запрещены. Русский язык не допускает, чтобы после начального *н* следовала согласная, как в словах НТО и НТР. Русский язык не допускает стечения пяти согласных подряд, как в словах ВЦСПС или РСФСР. Но это, так сказать, ограничения кода простого. Наш же язык недаром назван в заголовке этого очерка удивительным кодом. Несмотря на все свои правила, он ухитряется эти правила нарушать — и грамматические, и смысловые и, как вы сейчас убедились, фонетические (кстати сказать, не только в аббревиатурах, но и в иноязычных словах и именах, попавших в русский язык, также происходит нарушение кодовых правил: в наименовании народа Сибири *нганасаны* после *н* идет согласная; в слове *контрстратегия* шесть согласных следуют подряд; примеров же собственных имен, нарушающих правила фонетики русского языка, можно привести сколько угодно).

Еще одна удивительная черта нашего языка — это его способность к саморефлексии. Книга «Звуки и знаки» рассказывает о языке. Написана она также на языке. Об этой книге, повествующей о языке, можно говорить опять-таки на языке. Научные труды лингвистов, о которых рассказывает наша книга, также написаны на языке. И посвящены они анализу языка... Словом, мы можем строить целую иерархию различных уровней. Есть обычный человеческий язык, на котором можно

говорить просто, говорить о нем самом, говорить о том, как язык этот говорит о нем самом, и т. д.

И еще одна иерархия есть в языке, которой не обладает ни один из технических кодов, ни одна из других знаковых систем, что существуют в человеческом обществе. Любой знак системы дорожных указателей, шахматной нотации, азбуки Морзе или морской сигнализации флажками имеет определенное значение. А какое значение имеют звуки или буквы, из которых складываются слова? Никакого! Это не знаки, а только составные части знаков или, как говорят лингвисты, фигуры, из которых строится языковой знак.

В первом очерке мы приводили слова Ельмслева о том, что в языке с помощью горстки фигур может быть построен легион знаков. Но и эти фигуры имеют сложное строение, свою иерархию. «Атомы речи», фонема, как показывают исследования последних лет, строятся из набора элементарных различительных признаков, своего рода «элементарных частиц» языка. Фонемы, в свою очередь, образуют фигуры более высокого порядка — морфемы, то есть корни слов и служебные частицы, приставки, суффиксы и т. д. Да и слова не являются изолированными и совершенно самостоятельными единицами в отличие от знаков кода. Существуют тысячи словосочетаний вроде «круглый отличник» или «круглый невежа», фразеологизмы и непереводаемые буквально на другой язык идиоматические выражения вроде «взять быка за рога», «час от часу не легче», «держи карман шире», «приказал долго жить»... Наш язык — сложнейшая иерархическая система, причем элементы ее в отличие от технических кодов могут выступать на различных уровнях (вспомните пример с римлянами, поспорившими о том, кто скажет самую короткую речь или напишет самый короткий текст!).

Этот вывод подтверждается и анализом языка методами математической теории информации. Уже первые опыты по угадыванию показали, что информация распределяется в текстах неравномерно, какой бы язык или стиль ни был взят. «Начала слов несут максимумы информации, в то время как последние буквы слов и особенно следующие за ними пробелы оказываются либо мало информативными, либо вообще избыточными, — пишет Р. Г. Пиотровский в книге «Информационные изменения языка». — Квантовый характер распределения ста-

тистической информации связан, очевидно, с теми особенностями, которые характеризуют работу головного мозга человека в ходе переработки им лингвистического текста».

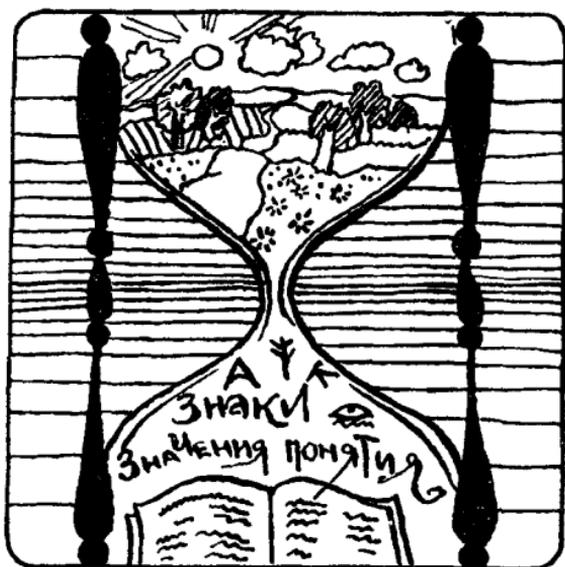
«Квантовое», зернистое строение имеют не только слова, но и словосочетания, да и вообще любые тексты. Связано это также с работой нашего мозга. И с тем, что любой текст состоит из элементарных фигур, образующих знаки-слова, которые в свою очередь, сочетаются в предложения, а из предложений строятся тексты... Так с помощью психологии, лингвистики, семиотики, математической теории информации начинают проясняться поистине удивительные свойства нашего уникального кода — языка.

Все системы знаков, которыми мы пользуемся, в сравнении с языком слишком жестки, вспомогательны, условны. В человеческом обществе язык был и остается основным средством передачи информации. И количество этой информации в наши дни начинает измеряться точными мерами, числами, понятными как человеческому, так и «электронному мозгу» вычислительных машин.

Однако язык способен не только передавать информацию о мире, который нас окружает. Он может и моделировать, своеобразно преломлять этот мир сквозь призму того или иного слова, выражения, текста, наконец, национального языка. И в этом — еще одно уникальное свойство человеческого языка, принципиальное отличие его от сигнализации животных, «языка машин» и любых технических кодов.

Соотношение языка, культуры и окружающего мира — вот предмет изучения этнолингвистики, науки, родившейся на стыке языкознания, этнографии и истории культуры. Об увлекательнейших проблемах этой дисциплины расскажет очерк

## МОДЕЛЬ МИРА



## ВОСКРЕШЕНИЕ ПРОШЛОГО

«Почему мы так говорим?» — эта рубрика часто встречается в популярных журналах. Рассказывает она о биографиях слов, которые порой бывают поистине удивительными. В самом деле: разве не интересно узнать, что немецкое *элефант* (слон) и русское *верблюд* происходят из одного и того же источника, вероятнее всего — древнеегипетского? Что наше *караул* происходит от тюркского *кара авыл*, то есть «охраняй аул»? А слово *акула* — из языка викингов?

Наука, изучающая происхождение слов, называется этимологией (от греческого *этимон* — истина и *логос* — учение, слово). Она помогает нам узнать, откуда и когда появилось в нашем языке то или иное слово, каким было его значение прежде (не так давно издательство «Прогресс» выпустило четырехтомный «Этимологический словарь русского языка», составленный Фасмером, в котором объяснено происхождение тысяч слов).

Но у этимологии есть задачи еще более интересные, важные и увлекательные. Происхождение слова, «история называния», не меньше характеризует человека и общество, в котором он жил, чем предмет, этим словом названный! Наука о «биографиях слов» помогает ученым раскрывать внутренний мир людей, умерших тысячи лет назад, мировоззрение «коллективов», казалось бы, бесследно исчезнувших, не оставивших никаких памятников — ни письменных, ни материальных. Подобно тому, как по костям умерших животных палеонтологи восстанавливают облик вымерших животных, так с помощью языка восстанавливается «модель мира», существовавшая в сознании наших доисторических предков.

Один всего пример. По-гречески *медведь* называется *арктос* (вспомните Арктику и то, что Полярная звезда входит в созвездие Малой Медведицы). По-латыни *медведь* звучит *урсус*, по-древнеиндийски — *ркшас*. Все эти родственные слова происходят от древнего индоевропейского названия медведя, которое звучит как *ркьтос*.

Русский язык — индоевропейский. Однако в русском, так же как и в родственных ему славянских языках, этого индоевропейского слова нет. Наше *медведь* буквально значит *медоед* — мед едящий, медв-едь.

В чем тут дело? Ведь такие слова, как *вода, нос, два, три, я, мать* и многие другие, остались похожими на древние индоевропейские. Они почти полностью идентичны в самых разных языках великой индоевропейской семьи (числительное *три* по-древнеиндийски звучит, как *три*, по-латышски — как *трис*, по-гречески — как *трэс*, по-немецки — как *драй*, по-хеттски — как *три*, на латыни — как *трэс* и т. д.). Почему же исчезло древнее название медведя?

Языковеды объяснили это. Предки славян, жившие в условиях первобытнообщинного строя, были суеверны. Они боялись называть медведя, хозяина, владыку дремучих лесов его собственным именем. И говоря о нем, страшное слово заменяли иносказаниями, намеками, позже это вошло в привычку, стало своеобразной традицией; не случайно герои охотничьих рассказов русских писателей, собираясь на медведя, избегали нередко называть его и говорили — хозяин, Он, топтыга... Одно из этих иносказаний закрепилось в русском языке — медоед, так появилось нынешнее — медведь. А не менее

суеверные предки германцев закрепили другое иносказание — бурый. Отсюда немецкое название медведя Bär (есть гипотеза, что и город Берлин обязан своим именем этому слову), английское bear и т. д.

Так язык помогает нам проникать во внутренний мир наших далеких предков вплоть до их суеверий и страхов перед дикими животными, хозяйничавшими в лесах.

### ОН, ОНА, ОНО...

Почему в нашем языке дом — это он, стена — она, а окно — оно? Ведь все неодушевленные предметы признаками пола не обладают. Да и существа одушевленные, если вдуматься, делятся у нас по какому-то странному принципу. У крыс есть и самки и самцы, однако мы говорим крыса и не называем самца крыс. То же самое можно сказать и о мышах, бабочках, мухах. Зато клоп — всегда он, так же как и комар, овод барс.

«На основе современного языка и современного мышления нельзя непосредственно уяснить, почему слова: потолок, сор, мор, сыр, жир, гроб, город, год и т. п. — мужского рода; стена, весна, плесень, плешь — женского рода, а поле, море, солнце, время, небо, лето — среднего рода, — писал академик В. В. Виноградов в своей монументальной монографии «Русский язык». — Самые мотивы распределения слов одного вещного круга (например, море, озеро, река, ручей, звезда, луна, солнце, месяц) по разным родам представляются непонятными. Так же неясно, почему живот мужского рода, а пузо или брюхо — среднего. Никто из говорящих на современном русском языке не осознает причины, почему из названий деревьев вяз, клен, ясень, дуб — мужского рода, а липа, осина, береза, сосна, ива, ветла, черемуха и др. — женского; или почему, например, кроме слова дерево (и растение), нет других русских обозначений деревьев среднего рода».

Категория рода, по мысли академика Виноградова, в русском языке является своего рода палеонтологическим отложением, живым ископаемым, пережитком древних представлений. Предки славян, следуя своей «модели мира», делили явления природы, предметы, живые существа на определенные категории или классы. Каж-

дому из них приписывался тот или иной род. В английском языке, восходящем, как и русский, к отдаленнейшему прапрапредку, индоевропейскому праязыку, категория рода исчезла совсем. А в других языках мира, наоборот, деление на классы намного превосходит привычное для нас, русских, деление имен существительных на одушевленные и неодушевленные, на мужской, женский и средний род. Вот, например, язык суахили. Свыше пятидесяти миллионов человек, живущих в странах Восточной Африки, говорит на нем. Существительные в суахили распределяются почти на два десятка различных классов. Причем каждый из них оформляется особым грамматическим показателем подобно тому, как оформляем мы соответствующими окончаниями существительные мужского, женского и среднего рода (только в суахили, в отличие от русского, оформителями служат частицы, прикрепляющиеся к началу, а не к концу слова).

Деление на классы в суахили отражает «модель мира», сложившуюся тысячелетия назад, давно забытую и сохранившуюся как реликт в языке.

Судите сами. Первый класс — это класс людей, второй — множественные существительные класса людей, третий — класс деревьев, растений, предметов, сделанных из дерева, а также названий частей тела. Особый класс существует для наименований округлых предметов и плодов. Новый класс — для жидкостей и существительных с абстрактным или собирательным значением. Еще один — для неодушевленных предметов, а также... названий лиц с физическими недостатками и названий языков. Для животных — новый класс, для предметов домашнего быта — тоже и т. д. Причем слепой человек отнесен к классу вещей, как и слова раб, рабыня и вообще названия лиц, так или иначе несамостоятельных!

Разумеется, никто из граждан Кении, Танзании и других восточноафриканских республик, где суахили является либо родным языком, либо языком-посредником, не считает калеку неодушевленным предметом, а свою руку или ногу — родственной дереву или предмету, из этого дерева сделанному (вспомним, что и у нас месяц — он, ночь — она, солнышко — оно, но только в сказках все эти слова персонифицируются в существа, наделенные признаками пола).

Классы существительных в суахили — это пережиток древнего мировоззрения. Гораздо ближе к первобытной модели мира языки племен и народов, живущих в условиях, близких тем, в которых жили наши прапрапредки и лишь в двадцатом столетии вступили или вступают в мир современной цивилизации. Таковы бушмены, кочующие по пустыне Калахари, индейские племена, обитающие в джунглях Амазонии, аборигены Австралии. Но наиболее яркие примеры мы находим в многочисленных языках острова Новая Гвинея.

В течение многих тысячелетий, а то и десятков тысяч лет коренное население Новой Гвинеи жило почти в полной изоляции от «большого мира», особенно в глубинах этого огромного острова. Даже контакты между отдельными племенами и деревнями были минимальными. Вполне понятно, что в языках жителей Новой Гвинеи, папуасов, первобытная «модель мира» выражена более явно. Первобытное мышление отражается в языке не как пережиток далеких эпох, а как непосредственная реальность, хотя в действительности она уже кое в чем изменилась.

Вот, например, папуасский язык аснат. В нем пять различных классов — и каких! Первый класс — это стоящие предметы, узкие и высокие (в том числе деревья и люди). Второй класс — предметы сидящие, столь же высокие, сколь и широкие (в том числе дома и женщины). Третий класс — предметы лежащие, широкие и низкие (упавшие деревья, пресмыкающиеся, а также солнце и луна, когда они только что встали из-за горизонта). Четвертый класс — плавающие предметы (рыбы, лодки и сами реки). Пятый — предметы летающие (птицы, насекомые, предметы, висящие наверху или лежащие в том месте, которое европейцы назвали бы антресолями, то есть выше обычного направления взгляда). Каждое существительное в языке аснат относится к тому или иному классу — считается либо стоящим, либо сидящим, либо лежащим, либо плавающим, либо летающим.

Такое деление кажется нам странным и необычным. Но оно подчиняется строгой логике. Для жителя непроходимых джунглей и мангровых болот очень важно знать, стоит, лежит, сидит, плавает или летает интересующий его объект. Вот почему в языке аснат, в зави-

сности от пространственного расположения, предмет попадает в тот или иной класс.

Впрочем, нет пужды обращаться к далеким экзотическим языкам. Возьмем современный немецкий и сравним его с древненемецким. *Tier* — так называются в современном немецком языке животные, звери, живые существа. Однако прежде этим словом обозначались только четвероногие, бегающие животные, причем только дикие.

Для всех летающих живых существ — от пчелы до птицы существовало слово *fogel* (ныне же *Vogel* означает птица). Все ползающие обозначались словом *Wurm* (современное значение этого слова — червь): и черви, и драконы, и змеи, и пауки...

Таким образом, не только жители далеких земель, но и европейцы членили мир по-особому с помощью языка. Кстати сказать, и в славянских языках можно найти такое членение.

Вспомним старинное русское слово *гад* — оно обозначало и змей, и жаб, и ядовитых насекомых, и мифических драконов, и обитателей морского дна («и гад морских подводный ход» — Пушкин).

Все это — пережитки древних представлений, своеобразные живые окаменелости, дошедшие до нас благодаря языку. Иногда ученым удается проследить, как образуются такие «живые окаменелости». В одном из языков Новой Каледонии слова *мать* и *отец*, *печень* и *сердце*, *потомки* и *жизнь* (личная) входят в две разные группы существительных. Почему? Современные новокаледонцы не ответят на этот вопрос, а историки и этнографы — могут. Дело в том, что в течение многих поколений новокаледонское общество сохраняло пережитки матриархата. Печень считалась символом личности, ее сутью.

Потомки, продолжающие род, почитались еще более, чем жизнь человека, давшего им жизнь. Все эти сложные религиозные представления и породили деление на *мать — печень — потомки*, с одной стороны, и *отец — сердце — личная жизнь* — с другой.

Структура новокаледонского общества изменилась, древняя религия забыта. Но язык сохранил в своей структуре «дела давно минувших дней».

## УДИВИТЕЛЬНЫЕ ЧИСЛИТЕЛЬНЫЕ

В некоторых книгах можно прочесть, что первобытные народы «не умеют считать до трех». Дескать, у них есть слова для числительного 1, для числительного 2, а числительное 3 означает уже много. Но по такой логике мы, русские, умеем считать лишь до десятка: ведь словами мы обозначаем числа от 1 до 10, а 11 уже одиннадцать, 12 — два на десять и т. д. На самом деле любой народ, на каком бы первобытном уровне культуры он не находился, владеет счетом. И австралиец, и бушмен, и папуас отлично знает всех своих соплеменников и родственников, отличит трех убитых животных от четырех. Но в языках мира действительно сохранились в виде живых ископаемых пережитки первобытного конкретного счета. Ведь не одно тысячелетие понадобилось человечеству, чтобы осознать число вообще, независимое от свойств предметов.

В десятках языков мира числительное 5 имеет родство со словом *пять, ладонь, рука*. В языке острова Пасхи и родственных ему языках, на которых говорят жители Океании и Мадагаскара, слово *рима*, или *лима*, означает и 5 и руку. А вот как интересно считают папуасы, говорящие на языке маринд. Числительное 1 обозначается словом *сакод*, числительное 2 — *ина*. Числительное 3 своего собственного ярлыка не имеет, оно передается словами *сакод-ина*, числительное 4 — *ина-ина* (то есть 3 — это 1—2, 4 — это 2—2). Дальше же начинается счет по пальцам рук и ног. То есть 5 — это уже не *ина-ина-сакод* (2—2—1), а большой палец руки, 6 — указательный и т. д. Пальцев, как известно, у нас 20. До двадцати и ведется счет, числа более двадцати именуются словом много.

Еще более интересно обозначаются числительные в языке телефол, на котором говорит около четырех тысяч человек на стыке границ Западного Ириана и молодой республики Папуа Новая Гвинея. Числительное 1 — это мизинец левой руки, 2 — безымянный палец левой руки, 3 — средний палец левой руки, 4 — указательный палец левой руки, 5 — большой палец левой руки, 6 — левое запястье, 7 — левое предплечье, 8 — левый локоть, 9 — левый бицепс, 10 — левое плечо, 11 — левая сторона шеи, 12 — левое ухо, 13 — левый глаз, 14 — нос. Затем происходит переход на другую сторону те-

ла. Числительное 15 — это другой глаз, 16 — другое ухо и т. д., вплоть до 27, обозначаемого словами мизинец правой руки.

Числительное 27, в свою очередь, берется за основу дальнейшего счета, который доводится до «носа», то есть до сочетания 27 и 14 (мизинец левой руки и нос). Оно имеет смысл «очень много» — дальше счет уже не ведется.

Лингвисты обнаружили в языках папуасов Новой Гвинеи двоичное, пятеричное (по пальцам рук) четверичное (счет по пальцам, но без большого пальца), шестеричное (шестерка обозначается словами шесть-один, дюжица — шесть-два и т. п.), двадцатисемиричное (как в языке телефол) счисление!

Встречается там и привычное нам десятеричное счисление, на основе которого построена система числительных русского языка. Кстати сказать, в нашем языке сохранились пережитки тех времен, когда какое-то большое число было синонимом очень много, больше не бывает. «Имя им легион» — говорим мы; или употребляем выражения «тьма-тьмушая», «тьма народа». И тьма, и легион в прежние времена были наименованиями определенных чисел. А вот еще один пример «живого ископаемого» в русском языке, связанного со счетом.

Почему мы говорим «две штуки» только о неодушевленных предметах? Это остаток первобытного конкретного счета в нашем языке. В других языках есть специальные числительные для подсчета различных предметов: для длинных одни, для круглых другие, для живых существ третьи и т. д.

Удивительную систему конкретных числительных обнаружил советский этнограф Е. А. Крейнович у нивхов, загадочных обитателей Сахалина и низовий Амура. «У них нет слова для обозначения абстрактного понятия «равный», но есть ряд слов для обозначения конкретных равенств. Нет у них и числительных для счисления абстрактных количеств, но зато есть примерно тридцать разрядов числительных для обозначения конкретных количеств», — указывает Крейнович и приводит длинный список таких числительных.

Тут есть числительные для подсчета предметов разной формы, отдельно мелких круглых (пуль, дробинки, яиц, икринок, кулаков, капель воды, топоров, бутылок), отдельно длинных предметов (деревьев, кустарников,

ребер, волос, кишок, дорог), отдельно плоских тонких предметов (листов бумаги, циновок, одеял, растений), отдельно парных предметов (глаз, щек, рук, лыж, весел, руковиц, серег).

При счете живых существ, семейств, поколений употребляются свои числительные, причем отдельно считаются люди и добрые духи, отдельно семейства, отдельно поколения, отдельно животные, рыбы, птицы, злые духи.

Для счета некоторых орудий лова рыбы и тюленей у нивхов опять-таки существуют свои особые числительные — для сетей; для неводов; для палок на острогу и т. д.

Особыми числительными ведется счет заготовленной рыбы и жердей для ее заготовки (для связок юколы для людей; для связок корюшки для людей; для связок корма для собак; для жердей для сушки юколы).

Есть еще специальные числительные при счете средств передвижения (отдельно лодок и отдельно нарт) и материалов (досок; прядей для веревок; связок травы для обуви).

Наконец, есть числительные, обозначающие меры: особо для маховых сажень, особо для пядей, особо для толщины сала медведей и тюленей, особо для дней в пути...

«Нетрудно заметить, что в группе числительных для счетов предметов разной формы отражена попытка человека провести классификацию предметного мира,— замечает Крейнович.— К какому же времени может быть отнесена эта классификация? Наличие топора в группе мелких круглых предметов подсказывает нам, что речь идет не о современном железном топоре, а об овальном топоре каменного века. Значит, классификация предметов по чисто внешнему признаку — форме, которая прослеживается в нивхских числительных, представляет, вероятно, одну из древнейших классификаций, созданных людьми каменного века».

### ГИПОТЕЗА УОРФА

Итак, язык помогает нам реконструировать «модель мира» наших далеких прапредков. В одних языках следов первобытного прошлого больше, в других меньше. Ведь и народы, носители этих языков, живут в разных

общественных условиях, имеют различный уровень культуры.

До сих пор речь шла о прошлом... Но где гарантия того, что и сейчас мы не воспринимаем, не моделируем мир сквозь свособразную призму нашего родного языка? Воспринимаем мы этот язык с младенческого возраста, он встраивается в наше сознание и подсознание. А затем мы считаем само собой разумеющимся то, что на самом деле велят нам законы родного языка.

Классический тому пример: семь цветов радуги. В русском языке каждый из цветов имеет собственный словесный ярлык. А вот в немецком или английском голубой и синий обозначаются одним и тем же словом. В языке одного из народов, живущих в африканской республике Либерия, наши семь цветов радуги обозначаются только лишь двумя словами: одно слово обозначает «холодные» тона (голубой, фиолетовый, синий), другое — «теплые» (красный, оранжевый, желтый, зеленый).

Не означает ли это, что и восприятие цвета идет через призму языка, через словесные ярлыки, которые в нем существуют? Вопросом этим пачали задаваться еще в середине прошлого века. Но свою полную и четкую формулировку он получил в трудах американского ученого Бенджамина Ли Уорфа.

«Я столкнулся с этой проблемой в области, обычно считающейся очень далекой от лингвистики,— писал Уорф.— Это произошло во время моей работы в обществе страхования от огня». Анализируя причины пожаров, Уорф обратил внимание на то, что многие несчастные случаи происходили... из-за слов. Например, рядом с тиглем для плавки свинца лежала грудка свинцового лома. Разве может гореть свинцовый лом? Никаких противопожарных мер не было принято. На самом же деле этот лом состоял из листов старых радиоконденсаторов, имевших парафиновые прокладки. Парафин загорелся, возник пожар. Вот и выходит, что причиной его стали слова, вернее, вера людей в то, что слова правильно называют вещи...

Уорф в свободное от работы время изучал письменность и культуру ацтеков и майя. Занятия эти разбудили в нем интерес к языкам коренных жителей Америки. На территории США и по сей день индейцы говорят на языках своих предков... Изучение же индейских языков

показало Уорфу, насколько различным может быть членение мира у различных народов.

Мы считаем само собой разумеющимся, что существительные обозначают предметы, а глаголы — действия. Однако на самом деле в окружающем нас мире, вечно меняющемся, нет отдельно предметов и действий. Что такое молния, волна, пульсация? Предметы или действия? Мы относим их к существительным, то есть к предметам. А в языке индейцев хопи, живущих в США, слова эти — глаголы. В языке же нутка, на котором говорят жители канадского острова Ванкувер, все слова показались бы нам глаголами. Это если мерять на аршин нашего родного языка. На самом же деле в языке нутка нет деления на предметы и действия, а есть единый взгляд на природу. Он-то и порождает один класс слов.

О доме на языке нутка можно сказать, что он стоит, но можно — и домит. Пламя может иметь место и может пламить. С помощью суффиксов и окончаний на языке нутка можно образовывать слова, которые придут слову *дом* различные оттенки длительности во времени: давно существующий дом, временный дом, будущий дом, дом, который раньше был, то, что начало быть домом, и т. д.

В языке индейцев хопи есть существительное, которое может относиться к любому летающему предмету или существу, за исключением птиц. Птицы же обозначаются другим существительным. «Можно сказать, — писал в одной из своих работ Уорф, — что первое существительное обозначает класс Л — П, то есть летающие минус птицы». И хопи называют одним и тем же словом и насекомое, и летчика, и самолет, но не птиц!

Даже такие, казалось бы, извечные и всеобщие представления, как пространство и время, получают разные языковые ярлыки. Как можно обойтись без глаголов в настоящем, прошедшем и будущем времени? Казалось бы, это невозможно. А вот индейцы хопи обходятся наклонениями. Утвердительное наклонение (*сообщаю о его приходе*) может относиться и к тем процессам, которые мы обозначили бы прошедшим или настоящим временем (*он пришел* или *он приходит*). Предположительное наклонение (*высказываю предположение о его приходе*) соответствует нашему *он придет*, или *он, наверно, приходит*, или *он, наверно, пришел*.

Мы измеряем время днями и годами, то есть существительными, хотя никаких предметов эти существительные не обозначают. У индейцев хопи это невысказано: существительные в их языке обозначают только настоящие предметы, физические тела. Вместо нашего «прошло два дня» индеец сказал бы: «В третий раз светает». Или, если обойтись без слова *раз* (и слегка насилуя русский язык), эту фразу с хопи можно было бы перевести как «третьеразно светает». Наше же русское «прошло два дня», если его буквально перевести на язык хопи, вызвало бы у индейца самый искренний смех: разве у дней есть ноги и дни ходят парами, взявшись за руки?

Но скорее всего, индеец хопи вообще бы не понял смысла фразы «прошло два дня». Ибо второй день для него — это не второй предмет, а тот же процесс, который прервался, а теперь возобновился (мы же не говорим «третий Иван Иванович приходит», если Иван Иванович ведет третий урок).

Языки разбивают окружающий мир по-разному с помощью слов и грамматических категорий. Не означает ли это, как предположил Уорф, что «каждый язык имеет свою метафизику»? И не была бы картина мироздания в физике Ньютона иной, если бы он говорил и думал не по-английски, а на языке хопи? (В одной из своих работ Уорф утверждал именно это.) Какова роль языка в восприятии, осознании, моделировании мира?

Общепризнанная заслуга Уорфа заключается прежде всего в том, что он показал на конкретных фактах важность роли словесного моделирования мира в различных языках; а вот его гипотеза вызвала бурную полемику. В Чикаго после смерти Уорфа был собран симпозиум, посвященный обсуждению его гипотезы. Приняли участие в нем лингвисты и логики, психологи и антропологи, этнографы и философы. «Язык и культура» — таков был главный вопрос симпозиума.

### РЕПОРТАЖ С ЛУНЫ

Определяет ли язык мировоззрение людей? Гипотеза Уорфа отвечала на этот вопрос положительно. Большинство современных ученых дает на этот вопрос иной ответ. Язык влияет на мышление, но только не на суть его, а на технику. Несмотря на различную технику языка (и, стало быть, детали техники мышления), любой

язык способен правильно передавать сообщения об окружающем нас мире.

Вот как рассуждал на симпозиуме в Чикаго известный американский языковед Джозеф Гринберг. Допустим, на Луну попадают два человека, говорящие на разных языках. Они оказываются совершенно в новой обстановке: мир Луны абсолютно не похож на наш, земной, здесь свои законы. Оба человека, побывавших на Луне, возвращаются на Землю и рассказывают о том, что они видели на чужой планете. Ведь, если следовать Уорфу, перед ними должны возникнуть два совершенно различных мира, две разные Луны.

Впрочем, не обязательно забираться в космическое пространство. В истории человечества есть немало примеров, сходных с «путешествием на Луну». Тысячу лет назад арабские путешественники посещали северные страны. Природа, нравы, обычаи норманских викингов были чужды арабам почти так же, как нам, землянам, чужд мир Луны. Арабский язык также не имеет ничего общего с языком норманнов. И все же описания жизни и быта викингов, сделанные на арабском языке, совпадают со скандинавскими хрониками: мы узнаем в них те же явления, события, города, горы. Язык по-разному «окрашивает» мир, но в конечном счете передает сообщения о действительности и передает их правильно.

Современные лингвисты сравнивают язык с системой координат. Перейти от одного языка к другому — это как бы перейти от одной геометрической системы отношений к другой. Окружающий мир, координаты которого дают языки, один и тот же. Различны лишь его отображения в языке.

Уорф совершенно правильно отметил, что в известных случаях язык может оказывать влияние на наше мышление (на его технику, а не на его существо, как уже подчеркивалось выше). Иной раз он может повлиять и на поведение людей: вспомним историю со «свинцовым ломом». Но Уорф прошел мимо другого, гораздо более важного факта: на мышление влияет прежде всего действительность, практический опыт людей, сама жизнь. И последнее слово принадлежит им, а не языку.

Несведущие люди не различают тонкостей, которые в жизни им обычно не нужны. Для простых смертных свинцовый лом гореть не может. Однако горький опыт с

пожарами научит разбираться в отличиях свинцового лома от свинцово-парафинового. И, если будет надо, то в язык войдет новое слово, с помощью которого можно разделить невоспламеняющийся свинцовый лом от горючего свинцово-парафинового. Мышление может, по вине ли языка или по любой другой причине, неправильно отражать мир. Но это возможно лишь до тех пор, пока практика и жизненный опыт не заставят людей уточнить «картину мира».

«От живого созерцания к абстрактному мышлению и от него к практике — таков диалектический путь познания истины, познания объективной реальности» (В. И. Ленин). Конечно, «итоги» этого познания закрепляются в языке, становятся смыслом слов, содержанием понятий, суждений, гипотез и теорий. Но никак не наоборот, как считал Уорф, преувеличивавший роль языка в процессе познания.

Но, как часто бывает в истории мысли, самые спорные взгляды оказываются самыми плодотворными. Недаром же один из наиболее строгих критиков гипотезы Уорфа сказал, что ошибки Уорфа гораздо интереснее избитых банальностей слишком осторожных ученых.

Гипотеза Уорфа (точнее — споры, ею вызванные) заставила заново пересмотреть многие, казалось бы, само собой разумеющиеся положения.

Споры вокруг гипотезы Уорфа, в которых участвовали философы, лингвисты, психологи, этнографы, социологи, носили отнюдь не отвлеченно философский характер. В последние годы был проделан целый ряд конкретных исследований, связанных со сложнейшим комплексом проблем, что кратко формулируется как «взаимоотношение действительности, мышления и языка». Особенно интересные результаты были получены при изучении наименования цветов — одного из краеугольных камней в гипотезе лингвистической относительности, как именуют еще гипотезу Уорфа.

### «ПАРАМЕТРЫ» ЦВЕТА

О том, что физически цвет представляет собой непрерывный континуум, мы узнаем из учебника физики. По данным физиологов, наш глаз может различать до десяти миллионов (!) цветовых оттенков... А что говорят данные лингвистики? Разумеется, ни в одном язы-

ке мира нет стольких словесных ярлыков для названия цветов и их оттенков. Мы обходимся несколькими десятками наименований. Причем в различных языках слова-ярлыки для одной и той же части спектра могут не совпадать. Например, часть спектра, обозначаемая в русском языке как *зеленый—синий—голубой—серый—коричневый*, в уэльском языке «сжата» до трех слов. Первое соответствует нашему *темно-зеленый*, второе — *светло-зеленый, голубой, синий, светло-серый*. А третье — *темно-серый и коричневый*.

Да и в одном и том же языке в различные эпохи спектр может члениться по-разному. В древнегреческом, в отличие от современного греческого языка, существовали слова-ярлыки для желто-зеленого и красного (*охрос*, отсюда наше — *охра*); для желто-зеленого и серо-коричневого (*хлорос*, отсюда наше — *хлор*) и для желтого, черного и темно-синего (*ктанеос*).

Чем же вызывается различное деление спектра в языках? Исследования последних лет показали, что огромную роль здесь играет... интенсивность солнечного света.

Как известно, солнце в тропиках стоит над горизонтом иначе, чем в наших, умеренных, а тем более — полярных широтах. Освещенность в тропиках больше, свет ярче. Чем дальше от экватора, тем легче воспринимать различия в цветовых тонах и труднее — в яркости цвета. Быть может, в языках жителей тропических широт существует своя особая терминология, обозначающая именно яркость, интенсивность освещения, а не градацию цветовых тонов?

Голландский ученый Г. ван Вейк проверил эту гипотезу на материале многих языков мира. Оказалось, что в большинстве языков народов, живущих в низких широтах, белый цвет и другие светлые тона обозначаются одним термином, черный и все темные — также одним общим словом. Причем это не группа цветов, как в европейских языках (так, темными мы считаем черный, фиолетовый, синий; светлыми — голубой, желтый, белый), а обозначения степени яркости — то, что в переводе на наш язык можно назвать темнотой и светлотой.

Между полюсами темноты (низкая яркость) и светлоты (очень большая яркость) находятся еще две категории: умеренная светлота (высокая яркость) и светлая темнота (умеренная яркость). Вот по этим катего-

риям яркости и распределяются привычные для нас наименования цветов.

Так, в языке хануно (остров Миндоро в Филиппинском архипелаге) есть четыре словесных ярлыка: светлота обозначается как *малагти* (сюда относятся все светлые тона); умеренная светлота — *малатуй* (зеленоватые тона); умеренная темнота — *марара* (красноватые тона) и темнота — *мабиру* (черный и все темные тона). В языке батак (остров Суматра) остается членение на четыре слова-ярлыка, хотя цвета распределяются по-иному: к светлоте относятся белый и светлые тона, к умеренной светлоте — светло-оранжевые и светло-коричневые, к светлой темноте — красные тона, оранжевый, розовый, коричневый; к темноте — темные тона, зеленый, коричневый, индиго.

«Хотя в языках народов экваториального пояса встречаются и системы обозначений по цветовому тону, статистически в большинстве языков этих широт, вне зависимости от расовой принадлежности говорящих на них, зрительные впечатления передаются яркостной номенклатурой», — констатируют ученые, занимающиеся этой проблемой.

При проверке гипотезы ван Вейка оказалось, что в терминах яркостей передаются цветовые впечатления и в языках чукчей, эвенков и других народов Крайнего Севера, удаленных от экватора на многие тысячи километров. Дело в том, что в Заполярье снежный покров создает огромную экранизирующую поверхность, и она делает освещение столь же интенсивным, как в тропиках. Вот почему чукчи называют все светлые тона словом *нилгыкын*, темные — *нивкын*, а красные, желтые и светло-коричневые — *ничелгыкен*; селькупы имеют словесные ярлыки для светлого (*сыры*), темного (*сэк*), красного (*няркы*) и желто-зеленого (*паталь*) и т. д.

Между тем многочисленные эксперименты показывают, что цветовое зрение у всех нормальных людей, независимо от их языка, расы, нации, культуры, одинаково (а дальтонизмом, например, европейцы страдают в два — четыре раза больше, чем аборигены Австралии). Но не только степень освещенности, не только «астрономические» причины влияют на цветовую номенклатуру языков. Названия цветов неразрывно связаны с окружающей природой, ее красками. Морское побережье и пустыня — вот два ландшафта жителей

африканской республики Либерия. Вот почему в языке басса, на котором говорят в Либерии, одно слово охватывает фиолетовый, синий и зеленый цвета, а другое — желтый, оранжевый и красный.

Еще большую роль играют не природные, а социальные условия. Аргентинские пастухи гаучо имеют две сотни названий для масти лошадей. В нашем русском языке у специалистов существуют также десятки разнообразнейших наименований масти. Есть тут и *буланный*, и *каурий*, и *сивый*, и *мышинный*, и *соловый*, и *пегий*, и *саврасый*, и *вороной*, и *гнедой* и т. д. Горожане этих слов, как правило, не знают. А если и знают, то затрудняются назвать, какой оттенок цвета имеет в виду тот или иной термин, обозначающий масть. Зато горожане, в особенности горожанки, перечисляют множество слов, обозначающих тот или иной оттенок тканей или губной помады, о которых представления не имеет крестьянин или животновод, превосходно знающий значения слов вроде *саврасый* или *каурий*.

Освещенность солнцем, природные условия, общественные потребности — все эти причины, вызывающие различное членение цветового спектра, так сказать, внешние по отношению к языку — все они свидетельствуют против гипотезы Уорфа.

Однако есть много фактов, которые говорят, что структура языка оказывает все же определенное влияние на моделирование мира в таком узком аспекте, как обозначения цветов. Например, во многих языках обозначения цветов конкретны, подобно тому, как конкретны наименования числительных, о чем мы уже рассказывали выше. На островах архипелага Бисмарка слово *коткот* обозначает ворону и черный цвет, слово *лулуба* — черную грязь болот в мангровых зарослях и соответствующий оттенок черного цвета, слово *угур* — обугленные листья бетеля, смешанные с маслом, и их цвет и т. д. (сравните наши прилагательные *вороной*, *агатовый*, *угольный*, *пепельный*, *смоляной*).

Во многих языках, помимо классификации цветов по яркости, проводится различие их и по другим признакам. Светлые тона, как правило, обозначают нечто положительное, темные — отрицательное (но если у нас цвет траура черный, то у китайцев — белый). В языке хануно, о котором мы уже говорили, одни цвета считаются влажными, другие — сухими. В языке австралий-

ских аборигенов аранта одно и то же слово может иметь значение мягкий, упругий, плод определенного дерева и зеленый или серый — в зависимости от контекста.

Иными словами, между языком, человеческим коллективом и миром, который люди моделируют посредством языка, существует своеобразная кибернетическая обратная связь.

Причем связь эта необычайно сложна. Не только конкретный язык, но и уровень развития культуры самого носителя языка, его социальное положение во многом определяют моделирование мира средствами этого языка.

### ОТ «ЭТНО» К «СОЦИО»

Итак, доминирующую роль в членении спектра на словесные ярлыки играет жизненная практика, уровень культуры, род занятий людей, а вовсе не их физиологическое восприятие. В языках многих народов Севера не различаются оттенки красного и зеленого цветов. Зато там есть десятки названий для оттенков белого цвета. Почему? Вовсе не потому что будто бы северяне не воспринимают красного и зеленого, а различают лишь вариации белого цвета.

В жизни северян, живущих в тундре, не так-то уж часто приходится сталкиваться с зеленью, цветами и т. п. Им гораздо важнее знать, в каком состоянии находится снег. Ведь от того, каким он будет, зависит и охота, и судьба оленьего стада. И в языке ненцев для обозначения понятия «снег» есть около сорока различных слов; в языке оленеводов чукчей есть отдельные слова для первого снега, который должен растаять, и для первого снега, который больше не будет таять; для мягкого снега и для плотного снега; для весеннего снега с проталинами и для мокрого, размякшего, тонкого снега; для мерзлого снега и для снега, уплотненного ветром после снегопада; для мягкого снега, легшего на плотный снег, и для снега, подтаявшего днем... Вполне понятно, что все это многообразие породило и обилие названий оттенков белого цвета, связанного со снегом.

В зависимости от потребностей общества находится, как правило, и число различных цветовых ярлыков, слов, обозначающих цвета. Так, в языке ненцев около трех десятков наименований цвета (причем три из них обозна-

чают оттенки серого цвета, вернее, его светлоту и темноту). Этого вполне достаточно для жителей суровой тундры, вплоть до Октября живших в условиях родового строя. В русском, английском, немецком, французском языках можно насчитать до сотни простых, состоящих из одного слова, наименований цвета.

Но ведь помимо простых, есть еще и составные названия, все эти темно-вишневые, светло-зеленые и тому подобные оттенки. В ненецком языке их очень мало. Зато в немецком насчитывается до полутысячи, а в английском, если учитывать терминологию портных, торговцев тканями, лошадьми и т. д., около четырех тысяч!

Разумеется, нельзя считать, что язык ненцев «хуже» немецкого или английского языка. Если возникнет необходимость, и в ненецком появится столь же богатая терминология, связанная с названиями цветов. Да и не только в ненецком, в любом языке мира лексика может стать сколь угодно богатой и разнообразной, если только в этом есть практическая нужда. Например, в языке исландцев, овцеводов и рыболовов насчитывается до полутора тысяч слов, связанных с погодой! Ибо капризная и переменчивая погода Исландии играет важную роль в хозяйстве острова. Вот почему даже ветер разной силы обозначается в исландском языке разными словами: силой в один балл — словом *андвари*, в два балла — *кул*, в три балла — *гола*, в четыре балла — *калди* и т. д.

И внутри одного языка можно обнаружить интереснейшее взаимоотношение между ярлыками, обозначениями цвета, и уровнем культуры людей, этим языком пользующихся. Любопытные данные были собраны в нашей стране в самом начале тридцатых годов. Группа ученых под руководством крупнейшего советского психолога Л. С. Выготского направилась в отдаленные кишлаки и на горные пастбища Узбекистана.

Тут уже началась перестройка социально-экономического уклада и культурного уровня жизни, но далеко еще не все население было вовлечено в этот процесс. На одном полюсе его были представители колхозного актива и студентки, на другом — женщины, всю жизнь прожившие в пределах ичкари, женской половины дома, и неграмотные крестьяне-единоличники. С различными социальными группами и были проведены эксперименты.

Известно, что в любом развитом языке есть четкий

и краткий набор слов, обозначающих абстрактные цветовые категории, вроде наших наименований частей спектра (красный, оранжевый, желтый и т. д.). Слова эти, как правило, теряют связь с конкретными, образными названиями. В то же самое время в любом языке есть слова, сохранившие свою цветовую конкретность: *молочный, лимонный, малиновый, вишневый, кофейный, бирюзовый* и т. п. Узбекистан — страна древней культуры; узбекский язык имеет богатую цветовую терминологию — как обобщенную, категориальную, так и конкретную, образную.

Какие же названия оттенков практически применялись различными группами испытуемых? Сохранялось ли у них одинаковое соотношение категориальных и образных названий? Этими вопросами в первую очередь и задались исследователи. Испытуемым предлагались мотки шерсти или шелка двадцати семи различных оттенков. Требовалось назвать эти цвета.

Люди, начавшие жить уже по-новому, давали обобщенные названия цветов, иногда уточняли их (темно-желтый, светло-синий). И лишь в редких случаях они приводили образные названия вроде *винный, фисташковый, табачный* цвет. По-иному на вопросы отвечали женщины ичкари. Образные, наглядные названия оттенков у них преобладали. Тут были цвета ириса, печени, розы, гороха, персика, гнилых зубов, испорченного хлопка, озера, неба, мака, воздуха, темного сахара, цветущего хлопка, помета свиньи и т. д. и т. п.

Затем испытуемым предложили разбить все оттенки на несколько групп. Колхозный актив, студентки, молодежь легко справились с этим заданием. Совершенно иная картина наблюдалась у женщин, имевших богатую практику вышивания, но всю жизнь проживших в затворничестве, не получивших никакого образования.

«Как правило, данная им инструкция — разбить предложенные оттенки на отдельные группы — вызывала у них полное недоумение и реплики «этого нельзя сделать», «здесь нет похожих, вместе их класть нельзя», «они совсем не похожи друг на друга», «это — как телячий помет, а это — как персик». Испытуемые этой группы начинали обычно прикладывать друг к другу отдельные моточки шерсти (шелка), пытались выяснить их оттенки, растерянно качали головой и — отказывались от выполнения задачи, — свидетельствует профес-

сор А. Р. Лурия в книге «Об историческом развитии познавательных процессов», вышедшей в 1974 году в издательстве «Наука». — Некоторые из испытуемых заменяли требуемую группировку по основному цвету раскладыванием в ряд оттенков по убывающей светлоте или насыщенности, в результате чего в один ряд вводились бледно-розовые, бледно-желтые, бледно-голубые оттенки или возникал единый ряд оттенков, переходящих друг в друга без видимых границ».

Все испытуемые пользовались одним и тем же узбекским языком. Но в зависимости от уровня их культуры и образа жизни происходило различное членение цветового спектра. Не язык, а общество оказалось главным «виновником» того, что спектр был по-разному расчленен словами-ярлыками!

### ЭТНОЛИНГВИСТИКА

Наука о языке все больше и больше связывается с пауками об обществе, в котором и для которого язык существует и развивается. Так, не зная истории петровской эпохи, нельзя понять, почему в русской морской терминологии так много голландских слов: *койка*, *каюта*, *ватерпас*, *зюйдвестка* и т. п. Не зная истории географических открытий, нельзя понять, почему в языке жителей африканского острова Мадагаскар встречаются арабские и французские слова, а сам язык — мальгашский — родственен не языкам близлежащей Африки, а многочисленным наречиям жителей островов Тихого океана. Не зная истории религии, нельзя понять, почему в монгольском языке буддийская терминология заимствована из санскрита, хотя сам буддизм пришел в Монголию из Тибета и языком богослужения был тибетский, а не санскрит. Некоторые чисто языковые факты нельзя понять без знания зоологии. Так, в языках коренных обитателей Бразилии, индейцев, а также в ряде языков других обитателей жарких стран нет слова *попугай*. Зато есть много слов для различных видов попугаев. Почему? Да потому, что слово *попугай* объединяет, по сути дела, разные и довольно-таки далекие друг от друга зоологические виды. Слово *попугай*, означающее попугая вообще, есть только в языках народов тех стран, где попугаи не водятся!

А вот другой зоологический пример. Наименования животных экзотических, как правило, попадают из языка коренных обитателей страны, где есть эти животные, в языки других народов неизменными. Таковы американские *опоссум*, *ягуар*, *пума*, *ассапан*, австралийские *кенгуру*, *эму*, *коала*, *динго* и многие другие.

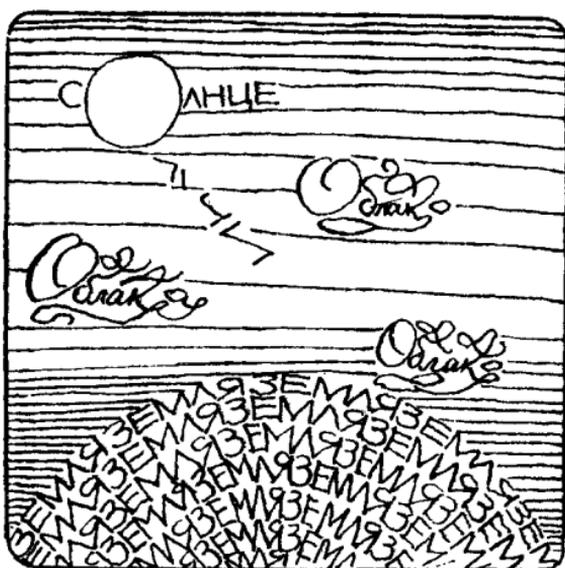
Еще большую роль играют данные этнографии, изучающей пищу, жилище, одежду, духовный мир различных народов и племен. Ведь лексика, словарный состав языка соответствуют тем общественным нуждам, которые есть у людей, этим языком пользующихся. Примеров тому можно привести великое множество, но мы ограничимся лишь одним. В эскимосском языке не было понятия и слова *воздух*. Зато явления, связанные с физическими свойствами воздуха, получали языковое обозначение: язык имел такие слова, как *пурга*, *дыхание*, *ветер* и многие другие. А для быстрой и точной пространственной ориентации в языке эскимосов существует более двух десятков указательных местоимений и свыше восьмидесяти слов, от них производных (не синонимов, а слов, обозначающих разные направления, разные точки в пространстве!). Ведь эскимосам, морским охотникам, ведущим коллективный промысел, необходимо предельно быстро и точно передавать «пространственную» информацию.

Мы недаром озаглавили предыдущую главку «От «этно» к «социо». Вероятно, вы и сами догадались, почему. Постепенно от этнографических факторов мы переходили к факторам социальным, от этнолингвистики — к социолингвистике. Однако очерк наш посвящен именно этнолингвистике. Проблемы же социолингвистики (или лингвосоциологии, как ее еще называют) охватывают очень большой круг вопросов. Чтобы рассказать о них подробно, понадобилась бы целая книга. И название ее было бы уже не «Звуки и знаки», а скорей всего, «Люди и знаки». Будем надеяться, что такая книга появится.

Как люди говорят? Каким образом с помощью звуков нам удается делиться нашими радостями и горем, мыслями и чувствами?

Эти вопросы не новы. Но лишь совсем недавно наука смогла установить, что же представляет собой одно из удивительнейших чудес на свете — наша обычная человеческая речь. Решающий вклад в это внесла фонология, дисциплина, роль которой в развитии гуманитарных наук сравнивают иногда с ролью ядерной физики в развитии естествознания... Почему? Об этом вы узнаете из очерка

## ФОРМУЛЫ ФОНЕМЫ



## ВРОЖДЕННЫЙ ИЛИ ОБЩЕСТВЕННЫЙ?

Египетский фараон Псамметих был любознателен. Однажды он пожелал узнать, какой из языков был первым на земле. Фараон повелел взять двух младенцев и отдал их на воспитание немому старику. Никто не смел подходить к ним, никто не смел обучать их языку.

Пусть речь их будет естественной, решил Псамметих. На каком языке заговорят подростки, такой язык и следует считать самым первым, самым ранним языком людей.

Легенда утверждает, что дети заговорили на языке фригийцев, народа, обитавшего когда-то на Балканах

и Малой Азии (кстати, сказать, язык фригийцев родствен современному армянскому языку). Они произнесли по-фригийски «бэкос, бэкос», то есть «овцы, овцы»...

Неизвестно, был ли проведен этот необычный эксперимент на самом деле. А если и был, то можно сказать с уверенностью, что фараона Псамметиха обманули. Кто-то ухитрился обучить детей говорить по-фригийски. Иначе быть не могло: сами дети никогда бы не научились говорить по-человечески без помощи людей.

Это наглядно подтверждают факты. Конечно, никто из ученых не пойдет на жестокий эксперимент, проделанный фараоном Псамметихом. Просто случалось так, что сама природа ставила естественные эксперименты на людях. Науке известны случаи, когда дети были вскормлены и воспитаны дикими животными — волками, леопардами, обезьянами, медведями и даже... овцой. И говорили эти люди-звери не по-фригийски и не по-русски, а по-звериному. Их горло издавало сигнальные крики тех животных, которые воспитали их: волчий вой, обезьяний визг, овечьё блеянье. Обучить их и говорить и жить по-человечески на все, как говорится, сто процентов было практически невозможно.

Ну а как обучается говорить нормальный ребенок, которого воспитывают люди, а не звери? Как же мы обучаемся говорить? Чтобы ответить на этот вопрос, ученые разных стран мира провели множество наблюдений, записывали лепетанье младенцев на магнитофон, тщательно анализировали эти записи. И вот какую картину они рисуют.

Уже в первые недели жизни младенец начинает реагировать на звуки. И не просто реагировать; он отличает неприятные, резкие и громкие звуки от приятных, мелодичных и тихих. Особенно же приятен для ребенка голос его матери (это не избитая фраза, а экспериментально установленный факт). Младенец выделяет голос матери из общего шума, он радуется, слыша его.

Благодаря способности не только говорить, издавать звуки, вроде классического «уа, уа», но и слушать, отличать один звук от другого, ребенок начинает учиться по-настоящему овладевать человеческой речью. Известно, что подражать звукам, в том числе и человеческой речи, умеют попугаи, скворцы, другие птицы. Ребенок также имитирует речь. Но, что самое главное, он не только подражает, он играет звуками речи. И такая

игра присуща только людям. По-иному она называется лепетом.

«Детский лепет» — зачастую говорим мы пренебрежительно о речи взрослых и этим обижаем детей, ибо в жизни ребенка лепет играет огромнейшую роль. Из этой звуковой массы выкристаллизовывается стройная система языка.

Лепетать младенец начинает с трех-четырёх месяцев, причем звуки, которые он издает, одинаковы у всех детей мира, как бы потом ни различались языки взрослых. Ученые нашли в детском лепете самые разнообразнейшие и сложные звуки: там есть и шипящие, и свистящие согласные, и гортанные звуки, которыми так богаты языки Кавказа, и даже щелкающие звуки, которые есть только в языках бушменов и готтентотов.

Почему же языки людей различны, если детский лепет младенцев всего мира один и тот же? Почему из лепета не развиться одному единому языку?

Потому, отвечает современная лингвистика, что человеческий язык не появляется сам собой, в процессе развития ребенка, а прививается обществом. Под влиянием взрослых детский лепет переходит в детскую речь. Те звуки, которых нет в родном языке, забываются, исчезают: ведь ребенок не слышит их от взрослых, которые учат его повторять только те звуки, которые есть в родном языке. Если бушмена в возрасте нескольких месяцев поместить в семью англичан, его родным языком станет английский. И, наоборот, маленький англичанин прекрасно усвоит язык бушменов, если его отдать на воспитание бушменам.

Кстати сказать, во многих языках банту, жителей Южной Африки, можно обнаружить щелкающие, «бушменские» звуки. Дело в том, что люди банту часто женились на бушменках. Многие согласные звуки, что есть в языках банту, отсутствуют в бушменских наречиях. Поэтому, обучаясь языку своих мужей, бушменские женщины заменяли эти звуки щелкающими звуками своих родных языков. Дети банту и бушменок усваивали два варианта произношения одного и того же слова: нормальный — от отца и щелкающий — от матери. Вот почему в языках банту многие слова имеют два варианта произношения: с щелкающим звуком и без него.

Однажды мы уже сравнивали язык с ситом, но там такое сравнение подчеркивало «частный случай». Сравнение с ситом более подходит ко всему языку как системе, отсеивающей в лепете ребенка ненужные, «неродные» звуки и оставляющей лишь необходимые, теми, которыми пользуются окружающие. Что же это за сито? Что такое «система языка» и для чего она нужна?

### АТОМЫ РЕЧИ

Возьмите самую обыкновенную телефонную или адресную книгу. Раскройте ее и выберите фамилию, допустим, Буров. Замените первую букву следующей буквой алфавита — В. Получите Вуров. Снова замените на букву Д — получите Дуров. Еще одна замена даст вам Гуров. Точно так же, пересобирая буквы алфавита, вы получите Журова, Зурова, Курова, Турова и т. д.

Что заставляет нас отличать Букова от Вурова? Или, скажем, знаменитого дрессировщика Дурова от эстрадного артиста Шурова, или ленинградского дравидолога Гурова? Только одна буква. Одна буква придает слову совсем другой смысл.

Буква, вернее, звук речи, который она передает, позволяет превратить Дурова в Шурова, Шурова в Гурова, Гурова в Турова и т. д. Звук меняет смысл слова... И не только в фамилиях. Например, слова *бочка* и *почка*, *кочка* и *точка*, *ночка* и *дочка*. Или *полка* и *полька*, *масса* и *касса*, *боль* и *голь*. Это свойство звука менять смысла слова получило название смыслоразличительной функции. А сами звуки, меняющие смысл, своеобразные атомы речи, получили название фонем. Фонемы — звуки языка, которые нельзя заменить другим, не изменив смысл самого слова.

Произношение фонем может варьироваться: в зависимости от голоса говорящего, его возраста, пола, эмоционального состояния и многих других причин. В слове *муму* звук *м* произносится с вытянутыми губами. Такое *м* и *м* без вытянутых губ в слове *мама* является одной и той же фонемой.

Почему? Попробуйте произнести *м* в слове *мама* так же, как мы произносим его в слове *муму*, вытянув губы. Смысл слова не изменится, несмотря на другое звучание фонемы *м*.

Но попробуйте-ка произнести вместо *м* — *п*. И сразу же мама превратится в папу! Или произнести *п* в слове *папа* звонко, тогда наш *папа* станет *баба*. Значит *м* и *п* — разные фонемы, так же как и *п* и *б*, несмотря на сходство их звучания.

Мы привели примеры из русского языка. В любом другом языке мира можно найти образцы подобного же рода превращений, связи звука и смысла. И, что самое интересное, очень часто случается так: в одном языке те же самые звуки являются вариантами одной фонемы, а другом — это разные атомы языка.

Возьмем наше *к*. Мы можем произнести его по-разному: то кратко, отрывисто, то с придыханием, будто запыхались, то обыкновенно. От этого слова *кол* или *кон* не превратятся в *гол* или *тон* и автор настоящей книги останется Кондратовым. Но попробуйте произнести *к* с придыханием в языке хинди. И сразу же получите иное слово. *Кар* на хинди значит *рука*. А вот то же слово, произнесенное с придыханием: *к'ар* — значит *осел*. В аварском языке, на котором говорят в Дагестане, на месте нашего *к* может оказаться четырнадцать (да, да, четырнадцать!) различных *к*. И все они — разные фонемы. Если перепутать их, получится не то слово.

Каждый язык по-своему членит мир звуков на фонемы. Каждый язык имеет свое особое сито, сквозь которое отсеиваются только нужные фонемы. Каждый язык имеет свои атомы, дробя на них непрерывный поток звуков, представляющий собой речь говорящего. Какими необычными на наш слух, на мерку нашего языка бывают эти атомы-фонемы!

Описание их напоминает порой описание экзотического растения или животного. Лингвисты, разумеется, выработали специальную терминологию, чтобы точно и адекватно описывать фонемы; разработан и специальный международный фонетический алфавит (МФА), состоящий примерно из сотни знаков для записи звуков любой речи. Однако и этих условных значков порой не хватает, когда ученые сталкиваются с удивительными фонемами, существующими в некоторых языках мира.

Со знаками МФА — Международного фонетического алфавита — знаком любой человек, изучающий иностранный язык. Именно этими знаками дается транскрипция слов (указание на то, как произносится слово). В каждом языке своя система фонем. Вполне понятно, что и для транскрипции применяется не весь набор знаков МФА, а только часть его. Однако когда лингвисты попробовали транскрибировать фонемы бушменского языка, оказалось, что знаки МФА к ним не подходят. Пришлось изобретать новые, ибо в наречиях бушменов есть совершенно особый тип звуков. Образуются они путем втягивания воздуха (а не выдыхания, как во всех остальных языках мира). Звуки эти называют щелкающими, или кликсами. А обозначаются они знаками: /, то есть черточкой; //, то есть двумя черточками; ///, то есть тремя черточками; +, то есть знак плюса; !, то есть восклицательный знак; ≠, то есть перечеркнутый знак равенства; ⊙, то есть кружок с точкой в центре.

Первый знак передает мягкий, «всасывающий» звук, который образует язык, прижатый к зубам; второй — звук, произведенный краем языка, коснувшегося неба и тотчас отдернутым назад и вниз; третий «квакающее» щелканье. Знак плюс передает щелканье «гортанное», восклицательный знак — звук, напоминающий хлопанье пробки. Перечеркнутый знак равенства обозначает звук, производимый кончиком языка, прижатым к корням верхних зубов («взрывное» щелканье, которое крайне трудно воспроизвести европейцу). Наконец, кружок с точкой в центре обозначает в записях лингвистов особый звук, подобный звуку поцелуя.

Но это еще не все. Щелкающий звук можно сделать звонким, а можно носовым, гнусавым. Его можно произнести с придыханием или без придыхания. И все эти вариации в бушменских наречиях создают новые фонемы. В отдельных бушменских наречиях число кликсов, щелкающих фонем доходит до трех десятков. А в одном из них на долю этих кликсов приходится около шестидесяти процентов всех звуков языка. И речь на таком языке создает удивительное впечатление какого-то потока щелканья, всасыванья, хлопанья пробки, кваканья, поцелуйных звуков и т. п.

Впрочем, странное впечатление может произвести на нас, привыкших к своим фонемам, любая иностранная речь, если ее атомы значительно отличаются от наших. Например, в языках Кавказа есть множество придыхательных, гортанных, «взрывных» согласных. В языке адыгейцев число согласных доходит до шестидесяти шести, в абхазских диалектах — до шестидесяти семи (а по мнению некоторых лингвистов, бзыбский диалект абхазского языка насчитывает до восьмидесяти фонем!), в абазинском — свыше семидесяти. Вполне понятно, что речь на подобных языках звучит очень непривычно для слуха тех, в чьем родном языке всего каких-нибудь два-три десятка согласных фонем.

«Странное впечатление производит этот язык на того, кто слышит его впервые! — писал один из первых русских кавказоведов П. К. Услар. — Об абхазском языке можно сказать, что он напоминает жужжание насекомых. Основа абхазского произношения состоит из сплетения самых разнообразных звуков, шипящих, свистящих, жужжащих, но разнородность их ускользает от непривычного слуха. Чтобы распутать это шаривари звуков, нет другого средства, как положиться на слух самих туземцев, которые всегда безошибочно укажут на звуки тождественные и на звуки разнородные».

Не только в языках бушменов или жителей Кавказа есть необычные на нашу мерку фонемы. Так, во многих языках Африки и Новой Гвинеи можно услышать так называемые преглоттализированные звуки. Вот, например, как произносится преглоттализированный звук *д*. Язык прижимается к деснам и одновременно смыкаются голосовые связки. Полость рта, таким образом «заперта». Вслед затем, по-прежнему держа рот закрытым, начинают раздвигать челюсти. В замкнутой полости рта возникает пониженное давление, голосовые связки начинают вибрировать, произнося этот звук. Потом язык отрывается от десен, и благодаря пониженному давлению во рту, туда попадает немного воздуха (наше описание заняло немало места, а звук, естественно, произносится очень быстро в разговорной речи африканца или папуаса).

В русском языке одно *а* — одна фонема. В латыни — две: *а* долгое и *а* краткое. В древнем языке индейцев майя три *а*, помимо долгого и краткого, есть еще *а*, прерываемое смычкой голосовых связок, так называемым

*сальтильо*, что по-испански означает *прыжок* (менее темпераментные датчане окрестили то же самое явление в своем родном языке словом *стед*, то есть *остановка*).

У нас две носовые фонемы, *н* и *м*. Казалось бы, какой еще носовой звук, кроме *м* и *н*, может быть? Однако в некоторых языках Новой Гвинеи есть шесть носовых согласных. И при этом есть еще носовые гласные, подобные тем, что существуют во французском языке. Зато в тех же языках *л* и *н* — это одна фонема; одной фонемой могут быть звуки *т* и *р*, казалось бы, не имеющие между собой ничего общего...

Мы могли бы еще долго рассказывать об удивительных звуках, существующих в различных языках мира. Но вам, пожалуй, стала ясна наша главная мысль: нельзя мерять чужой язык на свой аршин, нельзя примерять фонемы родной речи к фонемам чужой. У каждого языка свои координаты, свои атомы.

Меньше всего этих атомов в языках коренных обитателей Австралии и языках Полинезии. Так, в языке австралийского племени аранта насчитывают лишь три гласные и десяток согласных фонем. В языке гавайцев пять гласных и семь согласных. На другом полюсе стоят языки народов Кавказа. Мы уже говорили об абхазском, адыгейском, абазинском языках. Но рекордсменом здесь является язык убыхов (сто лет назад убыхи переселились из района нынешнего Сочи в Турцию и ныне почти полностью утратили свой родной язык). В языке убыхов насчитывается восемьдесят две согласные фонемы! А гласные — всего-навсего две, причем многие лингвисты считают, что в убыхском гласные не две, а одна — гласная *а* (вторая же гласная — неясно выраженный звук, представляющий вариант первой гласной).

На нашей планете несколько тысяч языков. Число фонем в них, таким образом, варьирует от двенадцати-пятнадцати до семидесяти-восьмидесяти. Вот из скольких атомов строится любой язык мира.

### ФОНЕМНАЯ «ТАБЛИЦА МЕНДЕЛЕЕВА»

Итак, существует несколько тысяч языков, каждый из которых обладает своей собственной системой, своими атомами, своим набором фонем. Нет ни одной фонемы, которая была бы общей для всех языков мира.

Что же получается в итоге? Выходит, что на земном шаре существуют сотни тысяч фонем (перемножьте тысячи языков на десятки различных атомов этих языков). А так как в диалектах одного и того же языка могут быть разные фонемы, то это число еще увеличится... Нельзя ли найти некую систему, что позволила бы как-то сгруппировать все это множество атомов языка? Найти своеобразную таблицу Менделеева для фонем?

Звуки речи могут быть бесконечно разнообразны. Произношение фонемы может варьироваться от одного говорящего к другому, один и тот же человек может произносить фонему по-разному, в зависимости от настроения, ситуации и т. п. И все-таки фонема объединяет эти звуки воедино, являясь, как выразился один из создателей фонологии Дэниэль Джоунз, «маленькой семьей звуков». Или, как говорил академик Л. В. Щерба, «звуковым типом», в то время как реально произносимые звуки, «являющиеся тем частным, в котором реализуется общее», то есть фонема, будут лишь оттенками этой фонемы.

Но если с помощью понятия фонемы удалось сгруппировать все эти «оттенки речи» в «маленькую семью звуков», то, быть может, удастся и сами фонемы объединить в общие группы, понимая, конечно, что не может быть универсальных фонем.

Этой мыслью языковеды задавались давно. И отдельные всеобщие законы — лингвистические универсалии — были обнаружены уже на заре фонологии. Например, в каждом языке мира есть деление фонем на гласные и согласные. Ну, а другие общие признаки, другие универсалии — есть ли они?

На этот вопрос ученые смогли ответить лишь недавно, во второй половине нашего столетия. Для этого пришлось проанализировать самые различные языки мира, самые разнообразные системы атомов речи. Причем анализ велся по двум направлениям. С одной стороны, наша речь порождается движением голосовых связок, языка, губ — словом, речевым аппаратом, который у человека развит, как ни у одного другого живого существа на планете. С другой стороны, наша речь — это колебания волн, непрерывный поток звуков.

Звуки эти можно записать на магнитофон, на осциллограф, а затем провести анализ записи. Иными словами, дать физическую их характеристику, опираясь на

данные акустики. Но и для «произносительной» характеристики фонем в наши дни разработана достаточно точная терминология. Недаром же в лингвистике существуют все эти глоттализированные, сонорные, фрикативные, абруптивные, назальные и т. п. звуки — в этих терминах описывается работа речевого аппарата человека, производящего тот или иной звук (вспомните наше описание преглоттализованного звука *ɖ* — сколько понадобилось слов, чтобы описать его произношение, в то время как термин «преглоттализированный» сразу же укажет специалисту, каким образом надо произнести этот звук).

Таким образом, фонемы можно описывать с двух сторон — со звуковой, акустической и речевой, артикулярной, произносительной. И затем, сделав такое описание, выявлять признаки их сходства и различия.

Мы уже говорили об одном различительном признаке, свойственном всем языкам мира: в них есть гласные и согласные фонемы. В очень многих языках существует противопоставление фонем по звонкости — глухости. Так различаются наши *б* и *п*, *д* и *т*, *с* и *з*. Но, например, в полинезийских языках или языке майя противопоставления по звонкости — глухости нет. Признак звонкости или глухости может быть описан как в терминах артикуляции, так и в терминах акустики.

Чем отличается русское *л* от *р*? Тем, что *л* мы произносим плавно, а *р* — прерывно. Противопоставление по прерывности — непрерывности есть и в других языках мира, хотя есть языки, где *р* и *л* одна, а не две фонемы. Таким образом, найден еще один различительный признак, хотя и не всеобщий.

В русском языке нет согласных, произношение которых сопровождалось бы смычкой гортани. Но такие согласные необычайно развиты в языках Кавказа, есть они во многих языках Новой Гвинеи и Африки. И фонемы в них противопоставляются по этому признаку.

Точно таким же образом были выявлены еще несколько различительных признаков. Гласные фонемы могут различаться низкой или высокой тональностью, бемольной или простой, дизонной или простой. Согласные могут быть носовыми или неносовыми, компактными или диффузными и т. д.

И, что самое поразительное, число этих различительных (или, как именуют их еще, дифференциальных) признаков оказалось совсем незначительным. Более

сотни тысяч фонем нескольких тысяч языков мира могут быть описаны в терминах всего-навсего дюжины различительных признаков. Двенадцать универсалий позволяют людям строить собственную систему атомов речи — фонем, отличную от языка другого народа, подобно тому, как из элементарных частиц строится многообразие атомов, из которых состоит наша Вселенная.

Фонему образно сравнивают с кружком, а различные признаки, по которым она отличается от других фонем в системе своего языка, с цветными квадратиками, что налагаются на этот кружок. Если фонема звонкая, то на кружок накладывается темно-синий квадрат, если глухая — светло-синий. Если она прерывная, накладывается темно-красный квадрат, если непрерывная — светло-красный и т. д. Каждая фонема будет представлять комбинацию, или, как говорят лингвисты, пучок различительных признаков. Или, если продолжать наше сравнение, каждая фонема будет представлена в виде кружка, покрытого комбинацией квадратов разного цвета. И число цветов, светлых и темных, которое необходимо нам, чтобы описать любую фонему любого языка, оказывается равным всего лишь двенадцати, хотя по своему богатству и разнообразию фонемы языков нашей планеты могут соперничать с цветовыми оттенками, которые различают наш глаз в видимой части спектра.

Вы, вероятно, заметили, что различительные признаки строятся на противопоставлениях, причем двоичных, или, говоря языком лингвистики, бинарных. Но собственно говоря, почему эти противопоставления бинарны? А скажем, не тернарны, то есть не троичны?

Вопрос, казалось бы, бессмысленный, вроде вопроса о том, почему колесо круглое или почему мы говорим *a*? Однако все обстоит далеко не так просто. Вспомните начало этого очерка. Мы говорили о том, как младенец начинает овладевать родным языком, как он просеивает сквозь сито фонем все многообразие звуков речи, которое порождает его лепет. Для детского ума двоичный принцип «да — нет», «хорошо — плохо», «белое — черное», «можно — нельзя» гораздо проще и доступнее, чем другие, более сложные вроде «да — может быть — нет»; «хорошо — посредством — плохо» и т. п. Конечно, ребенок нескольких месяцев от роду не думает ни об этих различительных признаках, о которых мы и сами узнали каких-нибудь два десятка лет назад, ни о фонемах. Но

подсознательно он улавливает, что баба отличается от папы звонкостью *б* и глухостью *п*, а носовое *м* превращает папу в маму.

## ДЕТИ И РОБОТЫ

— Девочка, как тебя зовут?

— Лита.

— Лита?

— Нет, Лита!

— Ах, Рита!

— Да, Лита...

Девочка не может произнести фонему *р*. Но она прекрасно знает, что *р* отличается от *л*, которое она приносит в имени «Рита» и других словах. Хотя, разумеется, не имеет никакого понятия о дифференциальном признаке плавность — прерывность, позволяющем нам различать *р* и *л*. И, что самое примечательное, стоит ребенку интуитивно постичь тот или иной признак, как он начинает использовать его сразу для нескольких звуков. Например, поняв, что баба отличается от папы (то есть противопоставление по глухости — звонкости), он начинает отличать и дам от там, зайку от сайки (ибо *д* и *т*, *з* и *с* также различаются признаком глухость — звонкость).

С помощью фонологии мы начинаем постигать, каким же образом наши дети овладевают азами языка. Этим азам кибернетики пытаются обучить и детище человеческого разума — ЭВМ. Мечта о разговоре человека с роботом насчитывает не одно столетие. Но только в нашем веке она стала на реальную почву...

Прежде всего, ученые решительно отказались от мысли обучать машину пословно, то есть вводить в ее электронную память эталоны звучания отдельных слов. Ведь слов в языке десятки и сотни тысяч. И пропорционально числу этих слов растут трудности, связанные с их распознаванием — легко научить отличать по звучанию два и пять, но не так-то просто различить два и льва или пядь и пять. Память ЭВМ предельно нагружается словами, а на выборку нужного слова из словаря отводится очень мало времени.

Может быть, научить машину распознавать не слова, а фонемы? Но мы-то различаем их по смыслу — бабу от папы, пять от пядь. Машина же смысла не пони-

мает. Кроме того, произношение фонем во многом зависит от возраста, пола, происхождения, образования говорящего. Интуитивно мы делаем поправку на это, мы распознаем фонемы, произносимые с искажениями, так как сравниваем их с эталонами фонем, что хранятся в нашем мозгу. Но ведь электронный мозг машины не имеет подобного рода эталонов!

Наиболее эффективным оказалось решение использовать различительные признаки, ту дюжину языковых универсалий, о которых мы говорили. Плюс еще универсалии, связанные с особенностями произношения людей — речь быстрая, медленная, плавная, резкая и т. д.

Сначала, принимая звуковые волны, то есть человеческую речь, машина делит ее на различительные признаки. Затем она сравнивает их с эталоном целого слова, которое хранится в ее электронной памяти. Но не в виде полной записи слова, а в виде последовательности различительных признаков, кодирующих его.

У современных вычислительных машин есть два вида памяти: оперативная память, небольшого объема, но быстродействующая, и память долговременная, значительно превосходящая первую в объеме и столь же значительно отстающая в скорости. Быстрая оперативная память может быть использована для переработки звуков речи в различительные признаки, а память большого объема — для хранения всего словаря.

Сначала эта модель была предложена из чисто лингвистических соображений. Затем исследования, проведенные в нашей стране под руководством Л. А. Чистович, показали, что распознавание речи человеком также происходит по сходному принципу. Чтобы понимать человеческую речь, машина должна, оказывается, делать это «по-человечески»!

В настоящее время в нашей стране и за рубежом создано немало машин, распознающих человеческую речь. И с каждым годом они совершенствуются. Как пример приведем электронно-вычислительную систему «ТРЕШОЛД-500», созданную в Англии. Практическое применение нашло уже более полутысячи систем «ТРЕШОЛД». Она используется для контроля качества продукции на конвейерах, для управления станками, для сортировки товаров, для опознания личности, для проверки багажа в аэропортах, в системах программированного обучения, для обслуживания парализованных больных в

больницах (выключить электроприбор, вызвать врача или медсестру и т. п.)...

Как же распознает речь эта система? Звуки, которые улавливает электронное «ухо» машины, разлагаются на составляющие элементы, преобразуются в двоичный код и поступают в память ЭВМ. Тут они группируются в пять семейств по тридцати двум фонетическим признакам (создатели «ТРЕШОЛДА» ограничились делением на согласные, гласные, короткие паузы, длинные паузы, взрывные звуки). Чтобы машина смогла определить еще и индивидуальные особенности голоса, слова произносятся многократно...

ЭВМ расчленяет длительность произношения слов на шестнадцать равных временных промежутков. Затем выявляет в каждом из них тридцать два фонетических признака. Слово преобразуется в числовой код по этим признакам. Наконец, после многократных повторов одного и того же слова машина выводит «усредненный код», то есть получает характеристики индивидуально-го голоса, понятные ее электронному мозгу.

Общая емкость словаря системы «ТРЕШОЛД» — до двухсот двадцати слов. Словарь записывается на магнитной ленте, которая хранится в библиотеке словарей системы. Естественно, что словарь составляется для каждого человека и фиксирует особенности голоса с их индивидуальным тембром, окраской и прочими особенностями. Причем машина различает слова, хранящиеся в ее памяти, независимо от того, в каком настроении мы их произнесем — со страхом, радостью, болью и т. д.

## У КОЛЫБЕЛИ ЯЗЫКА

Конечно, система «ТРЕШОЛД» — явление не уникальное. Например, в университете Карнеги — Меллона, в США, разработана экспериментальная система «захвата речи», которая позволяет распознавать до девяноста пяти процентов фраз. Причем произносит эти фразы не один человек, а пять, три мужчины и две женщины, а словарь содержит более тысячи слов.

Работы по «захвату речи» ведутся в Национальном исследовательском центре дальней связи во Франции, ведут их также советские ученые — в Москве, Ленинграде, Тбилиси, Киеве... Рассказ об этих работах занял бы много времени, к тому же они все-таки относятся

более к технике, чем к лингвистике, основной теме нашей книги.

Вам, пожалуй, стала ясна огромная роль фонологии в попытках научить машину говорить «по-человечески». Еще более важную роль может сыграть эта дисциплина в понимании того, каким же образом сигнальная система наших прапрапрапредков, приматов, превратилась в человеческую речь.

Человеческие языки в среднем имеют тридцать-сорок атомов, фонем, из которых строятся молекулы слов. Изучая системы сигнализации наших ближайших родственников по древу жизни — приматов, ученые обнаружили, что число осмысленных сигналов, каждый из которых соотнесен с определенной ситуацией, находится в пределах двадцати-сорока знаков. Столько их у шимпанзе, у прославившихся на весь мир диких обезьян Японии, у других приматов.

Случайно ли это совпадение? По всей вероятности, нет, не случайно. Видимо, и у наших прапрапрапредков первоначально существовала сигнализация, состоящая из нескольких десятков знаков. Анализ черепов питекантропов, обезьянолюдей, пещерных людей, неандертальцев, показывает, что у них не было речи в нашем понимании, то есть человеческой членораздельной речи, представляющей сложную иерархию разных уровней — фонем, морфем, слов, предложений. По всей видимости, у них была примерно такая же примитивная сигнальная система, как и у человекообразных обезьян.

Трудовая деятельность вызвала необходимость в новых словах-сигналах. И не в одном-двух, а десятках, сотнях, тысячах. Остальным приматам достаточно было тридцати-сорока сигналов, чтобы выразить тревогу, удовольствие, призыв и тому подобные примитивные «понятия» и чувства. «Человеку разумному» этого было недостаточно.

Казалось бы, самый простой путь — увеличивать число сигналов-знаков, наращивать словарь. Однако это потребовало бы колоссального объема памяти, причем, если говорить языком кибернетики, памяти оперативной, быстродействующей. А ее объем, как известно, не может быть большим, иначе потеряется быстрота...

И тогда эволюция повела наших предков иным путем — тем самым, что ныне в какой-то мере имитируют современные ученые, когда обучают ЭВМ понимать че-

ловеческую речь. Нерасторжимые в прежней системе — в системе сигнализации приматов — сигналы-знаки стали делиться на элементарные различительные признаки. Из них стали формироваться фонемы — уже не простые «знаки ситуации», а единицы языка, служащие для образования единиц более высокого порядка — морфем, слов, а затем и предложений.

Вероятно, в недалеком будущем лингвисты в содружестве с антропологами, психологами, кибернетиками, социологами смогут показать наглядно, в деталях, как происходило превращение сигнальной системы человекообразных обезьян в язык наших предков... Как трудовая деятельность вызвала потребность в новых сигналах... Как вместе с ростом словаря возрастала и емкость памяти... Как трудовые навыки и увеличение объема словаря способствовали эволюции мозга... Как вместе с развитием мозга шло совершенствование речевого аппарата... Как законы языка закреплялись в сознании отдельных членов общества и всего общества...

Одним словом, как родился и развивался наш человеческий язык, справедливо именуемый чудом.

### **ЗООЛИНГВИСТИКА**

Фонология находит еще одно интереснейшее применение. С помощью методики, разработанной в фонологии, некоторые исследователи пробуют описывать сигнализацию у дельфинов, обезьян и других животных.

Советский ученый Н. И. Жинкин, специалист по физиологии и психологии речи, проделал обстоятельную работу, посвященную звуковой сигнальной системе обезьян — гамадрилов.

Работа эта была проведена по всем правилам современной лингвистики. «Звуковое» измерение криков обезьян было проделано с помощью осциллографа. Спектрограммы позволили Жинкину произвести «микроскопический анализ» звуков. «Речевое» измерение удалось провести с помощью рентгеноскопа. Он точно зафиксировал артикуляционные движения, которые делало горло обезьян при «разговоре».

И, наконец, полученные данные были обработаны согласно теории фонологических различительных признаков. Оказалось, что сигналы-«слова» гамадрилов составлены по меньшей мере из десяти элементарных

звук. Например, тихий и довольно сложный по звуковому составу сигнал удовольствия, который очень приблизительно можно передать как *хон*, где *х* — нечто похожее на придыхание, а *он* — ясно слышимое *о* с носовым резонансом — состоит из трех элементарных звуков. В принципе, пользуясь этими элементами, можно построить около тысячи сигналов-«слов». Гамадрилы же пользуются по самой щедрой оценке, всего лишь сорока сигнальными знаками. Больше им не требуется: ведь жизнь обезьяньего стада неизмеримо проще, чем жизнь человеческого коллектива. Это принципиально разные явления и принципиально различаются языки людей и сигнальные системы животных.

Порой сложное, «двухэтажное», строение имеют системы сигнализации и других животных, а не только высокоразвитых обезьян. Так, у курицы общий сигнал тревоги разделяется на четыре различных сигнала: опасность близко, опасность далеко, опасность — человек и опасность — коршун. Сигнал призыва, повторенный дважды, означает категорический приказ. Всего в «курином языке» около десятка знаков-кирпичиков, из которых слагается два-три десятка различных сигналов. А ведь в принципе из них можно было образовать сотни новых сигналов! Однако жизнь курятника еще более проста, чем жизнь стада гамадрилов. И поэтому система сигнализации у кур обходится двумя десятками «слов».

Звуковые сигналы издают львы и морские коньки, саранча и дельфины, да и вообще «разговаривают», обмениваются сигнальными звуками тысячи видов живых существ, в том числе рыбы, когда-то считавшиеся воплощением молчаливости. «Нем как рыба», — гласит пословица. Однако приборы, опущенные под воду, показывают, что тут идет беспрестанная болтовня, которую ведут обитатели подводного царства, обмениваясь звуками. Но за редкими исключениями, «рыбий язык» — ультразвуковые сигналы призыва, тревоги и т. п. — наше ухо не воспринимает. Так что пословица «нем как рыба» для нас, людей, остается в силе.

Ученые всего мира терпеливо и настойчиво собирают коллекцию звуков, которые издают тигры и лягушки, насекомые и птицы — словом, все живые существа нашей планеты. С помощью чутких микрофонов на магнитофонную пленку записывается стрекотание кузнечи-

ка и пенье цикады, писк комара (это тоже сигнал!) и щебетание птиц. Записи анализируются, делаются спектрограммы, выявляются отдельные сигналы в непрерывном потоке звуков.

### ВСЛЕД ЗА ФОНОЛОГИЕЙ

Язык справедливо называют системой, структурой, кодом, каким бы уникальным ни был этот код, какой бы сложной ни была эта структура, какую бы иерархию ни имела система. Если язык — система, если его нижний этаж, фонемы, подчиняются законам универсалий, то, быть может, и другие, более высокие этажи имеют подобные же законы? Нельзя ли найти различительные признаки в грамматике?

Например, такие категории, как род или число, невольно рождают мысль о своей всеобщности, универсальности. Род может быть мужским, женским или средним (или категория рода может вообще отсутствовать, как в английском языке). Число — единственным или множественным (или опять-таки отсутствовать, как в китайском языке). Разве не соблазнительно произвести классификацию языков мира по этим признакам?

Но когда лингвисты обратились к языкам Африки, Океании, Австралии, языкам индейцев Америки, а также ко многим языкам Азии, неведомым классическому языкознанию, они обнаружили в них явления, не подпадающие под привычные для нас мерки. Во многих языках Новой Гвинеи и Тропической Африки, как и в английском языке, категории рода нет. Зато есть, как мы говорили, особые классы существительных. В африканском языке суахили таких классов два десятка, в папуасском языке насиои — свыше сорока (класс мужчин, женщин, больших животных, птиц, рыб, плодов, дней, садов, групп людей, домов, пальцев, пакетов, вязанок, кусков и т. д. и т. п.). Порой один и тот же корень, соединяясь с тем или иным грамматическим показателем класса, приобретает разные значения. В языке племени байинг, живущем на острове Новая Британия, *мунг-ар* означает бревно, *мунг-инг* — кусок дерева, а *мунг-ит* — лучину, щепку.

Да и в языках, где есть категория рода, деление на роды бывает весьма необычным. В папуасском языке абелам есть «живой» род, куда зачислены люди, круп-

ные животные и части тела, и «неживой», к которому относятся вещи, насекомые, птицы и мелкие звери. В другом папуасском языке, монумбо, есть мужской, женский, средний, детский и смешанный род. А в третьем папуасском языке, маринд, слово *анем* означает человек, мужчина, слово *анум* — женщина, а слово *аним* — люди. Вот и разберитесь, где здесь категория рода, а где — категория числа!

Впрочем, и категория числа также весьма непроста. Есть языки, где помимо единственного и множественного числа, есть еще и двойственное число, и даже тройственное. В папуасском языке ава *иян* — это собака, *иятаре* — две собаки, *иятаро* — три собаки, *иямари* — много собак. В бушменском языке существует не менее шести способов образования множественного числа.

Как видите, не так-то просто находить универсальные признаки в грамматике языков. И все-таки лингвисты их настойчиво ищут. Должны же они в конце концов быть, эти универсалии — ведь недаром то, что сказано на одном языке, может быть переведено на другой, какими бы ни были различия в фонетике и грамматике. Первые опыты в этом направлении уже были сделаны.

Луи Ельмслев, датский лингвист, определил, что теоретически в языке может быть двести шестнадцать различных падежей. В реальных языках мира это число колеблется от нуля (в китайском) до пятидесяти двух (в табасаранском, одном из языков Дагестана). В древнейшем индоевропейском языке, прадедушке нашего русского, английского, греческого, персидского и многих других языков, родственных между собой, было восемь падежей. В старославянском языке было семь падежей, а в нынешнем русском их всего шесть (исчеззательный падеж, следы его остались в обращениях *отче, старче, боже*).

Нельзя ли попробовать описать падежную систему на основе различительных признаков? Разумеется, признаки эти уже будут не фонетические вроде звонкость — глухость, а смысловые. Ведь анализировать мы будем не звуки языка, а слова. Но сам принцип противопоставления, принцип «да — нет», остается.

Такой анализ был проделан Романом Осиповичем Якобсоном, которого справедливо называют наряду с замечательным русским ученым Николаем Сергеевичем Трубецким «отцом» современной фонологии. Проиллю-

стрируем эту работу на материале знакомого всем читателям нашей книги русского языка.

Начнем с именительного падежа. Чем он отличается от остальных? Прежде всего тем, что он как бы нейтрален, слово в именительном падеже является базисным, основным, исходным для остальных падежей. Винительный отличается от именительного тем, что действие в нем направлено на объект. Творительный отличается от именительного тем, что вовлекает в соучастие в действие и по отношению к главному действию он находится как бы на периферии; родительный — тем, что не полностью охватывает действие по объему...

Получаем различительные признаки: направленность — ненаправленность, периферийность — непериферийность, объемность — необъемность... Оказывается, их вполне достаточно, чтобы описать систему русских падежей в терминах различительных признаков. Каждый падеж, подобно фонеме, будет пучком этих признаков. Вот такая табличка у нас получится, если мы обозначим присутствие признака знаком плюс, а отсутствие — знаком минус.

Именительный: — — — (ненаправленный, непериферийный, необъемный)

Родительный: — — + (ненаправленный, непериферийный, объемный)

Дательный: + + — (направленный, периферийный, необъемный)

Винительный: + — — (направленный, непериферийный, необъемный)

Творительный: — + — (ненаправленный, периферийный, необъемный)

Предложный: — + + (ненаправленный, периферийный, объемный)

Схема эта применима и к другим славянским и индоевропейским языкам. Разумеется, для того, чтобы описать падежную систему таких языков, как табасаранский с его пятьюдесятью двумя падежами, потребуется введение новых различительных признаков. Но дело не в числе этих признаков, а в самом методе. Принципы фонологии могут применяться не только к фонемам, но и более высоким «этажам» языка!

Самое замечательное в различительных признаках падежей русского языка не то, что они описывают падежи, а то, что они могут быть применены и к другим грамматическим категориям. Как подчеркивал Якобсон, признак направленности может характеризовать еще признаки противопоставления глагола в залогах (действительный и страдательный). Признак объемности связан с понятием предела, и здесь мы видим связь с делением русских глаголов на совершенные и несовершенные. Так, отрицание одинаково воздействует на этот признак, касается ли это существительного в объемном родительном падеже или глагола несовершенного вида (возьми книгу — не бери книги).

Таким образом, открывается возможность анализировать не отдельные фрагменты грамматики вроде падежа или рода имен существительных, но и грамматику языка в целом. Причем в связи ее со значением, со смыслом слов.

Может быть, и для лексики удастся отыскать набор различительных признаков? Образуют ли значения слов систему, подобную той, что мы находим в фоиегической структуре языка?

Ответить на этот вопрос не так-то просто. Число фонем да и грамматических форм ограничено. Слов же в любом языке очень много. И, что самое главное, для них нет никаких количественных ограничений. Одни слова исчезают, другие появляются. «Поэтому при первом рассмотрении словарь представляется отрицанием понятия состояния, устойчивости, синхронии, структуры,— пишет Луи Ельмслев в статье, названной «Можно ли считать, что значения слов образуют структуру?» — Кажется, будто в словаре царят каприз и произвол и что поэтому словарь — это противоположность структуры».

Тем не менее и Ельмслев, и многие другие современные лингвисты полагают, что ответ на вопрос, заданный им в заголовке статьи, должен быть положительным.

Количество слов в словаре неограниченно. Это, говорят математики, открытое множество, в отличие от фонем или категорий грамматики, образующих множество закрытое. Но внутри словаря можно найти закрытые

множества, или классы. К ним относятся термины родства или обозначения цветов. Есть и совсем маленькие классы, состоящие всего из двух членов, например, производные прилагательные вроде *большой — маленький, длинный — короткий, красивый — безобразный, горячий — холодный* и т. п. Как справедливо замечает Ельмслев, структурное описание возможно лишь при условии, что открытые классы удастся свести к закрытым. Для этого необходимо проделать огромную по объему работу.

Такая работа проводится в различных странах мира. С одной стороны, лингвисты стараются разбить все многообразие лексики на отдельные семантические, то есть смысловые «поля». С другой стороны, делаются попытки описать то или иное конкретное семантическое «поле» в терминах различительных признаков по принципу «да — нет». «Двоичный набор признаков для классификации в последнее время строится для наиболее употребительных слов языка, не связанных с обозначением конкретных вещей, — констатирует советский ученый Вячеслав Всеволодович Иванов. — При изучении небольших подсистем слов с абстрактными значениями выделяются двоичные классификационные признаки, частично совпадающие не только с универсальными грамматическими признаками, но и теми признаками, которые оказываются существенными и для исследования мифологических и ритуальных систем».

В 1941 году, перед самой войной, советский этнограф А. М. Золотарев закончил монументальный труд «Дуальная организация первобытных народов и происхождение дуалистических космогоний». Безвременная кончина оборвала жизнь ученого, и труд его увидел свет лишь в 1964 году. И тем не менее книга его не устарела. Наоборот, она оказалась в русле современных идей, она дала уникальный материал для анализа с помощью различительных признаков не только мифологии первобытных народов, но и самой структуры общества, эту мифологию породившего.

Золотарев обработал тысячи источников и собрал данные о двоичной, дуальной организации самых разнообразных племен и народов. Столь же тщательно собрал он различные мифы и легенды, повествующие о близнецах, о борьбе двух начал в мире. Следы этой

дуальности, как показал Золотарев, можно найти не только у народов Австралии и Океании, Африки и Азии, Северной и Центральной Америки, но и в античном мире и на Древнем Востоке.

В обрядах и ритуалах самых разных народов мира можно найти ту же дуальность, те же противопоставления «да — нет». Достаточно назвать противопоставления *небесный — земной, святой — грешный, чистый — нечистый, божеский — человеческий*, которые есть в любой религии мира. Источник этого, как показывают материалы Золотарева, — «двоичная» структура общества, деление его на две фратрии, группы и т. п. А это значит, что методы фонологии могут быть применены не только в лингвистике, но и этнографии, истории религии, мифологии, истории первобытного общества.

Но, быть может, и в изучении современного общества эти методы также найдут применение? Некоторые ученые полагают, что и социология может почерпнуть для себя много ценного в фонологии. И советские, и зарубежные ученые не раз высказывали мысль о том, что член коллектива может рассматриваться как «пучок различительных признаков», программ поведения и языковых программ, которые существуют в этом коллективе.

И, естественно, чем меньше этих программ, тем беднее личность, тем менее развито общество. Сравните число каналов связи, в которые включен современный образованный человек с числом каналов, в которые включен бушмен пустыни Калахари или абориген Австралии...

...От анализа фонем, этих атомов языка, от поиска различительных признаков, своеобразных элементарных частиц, составляющих эти атомы, мы перешли к сложнейшим вопросам, связанным с человеком и обществом, в котором он живет. Иными словами, к социолингвистике. А она, разумеется, слишком сложна, интересна, серьезна, чтобы ограничиться рассказом в несколько страниц, да и одного очерка будет мало, здесь нужна целая книга.

Нет, очевидно, не зря роль фонологии в развитии гуманитарных наук сравнивают с ролью ядерной физики в развитии естествознания.

Фонология дает идеи другим наукам и находится в авангарде науки о языке.

## ЧТО БУДЕТ ДАЛЬШЕ?

Термин фонема был предложен в конце прошлого столетия замечательным русским ученым Иваном Александровичем Бодуэном де Куртенэ. Не удивляйтесь странной фамилии: в жилах Иваиа Александровича текла кровь крестоносца, провозглашенного в начале XII века «иерусалимским королем». В мемуарах Виктора Шкловского «Жили-были» Бодуэну де Куртенэ посвящены прекрасные строки как человеку и как ученому, и мы не будем их повторять. В иашей стране в двух томах изданы избранные труды Бодуэна де Куртенэ. Имя его пользуется почетом и уважением во всем мире — и прежде всего как основоположника фонологии.

Правда, в начале нашего столетия только русская школа Бодуэна де Куртенэ, включая его лучшего ученика Л. В. Щербу, принимала учение о фонеме как о своеобразном атоме языка. Лишь после первой мировой войны это учение получает отклик за рубежом и дальнейшее развитие.

Самый ценный вклад в современную фонологию внесли профессора Н. С. Трубецкой и Р. О. Якобсон. Учение о фонеме начинает распространяться во всем мире. После второй мировой войны, говоря словами известного советского лингвиста П. С. Кузнецова, «уже почти не остается ученых, которые бы считали нецелесообразным включение в том или ином виде понятия фонемы в число основных лингвистических понятий».

Сам термин фонема служит образцом для построения других моделей языка, изучающих не звуки, а другие, более высокие и сложные уровни его иерархии. Бодуэн де Куртенэ ввел термин морфема, образовав его от греческого слова морфе, то есть форма. Морфемы — это своеобразные атомы значения, минимальные значащие единицы языка (корни слов, приставки, окончания, суффиксы). Вслед за фонемой и морфемой появляются термины сема, семема, семантема. Ими обозначают атомы смысла (вспомните семантические множители, о которых рассказывал очерк «МП, ЯП, ИЯ»). В научной литературе наших дней можно встретить термины сонема, лексема, интонема, графема, граммема, кинема, тонема, номема, пиктема, релатема и многие, многие другие *емы*, что образованы по аналогии со словом «фонема». Только служат они для описания других элемен-

тарных единиц языков и систем знаков: атомов грамматики, лексики, графики, интонации, жестикюляции, тона и т. д.

А сама фонология, дав толчок к развитию самых разных областей языкознания, почилa на лаврах? Разумеется, нет. Здесь еще много спорного и нерешенного, начиная с описания конкретных языков с помощью различительных признаков и кончая самими различительными признаками, их универсальностью.

Ни в одном из языков мира, которые описывались с помощью различительных признаков, не встречаются вместе все эти признаки. Например, система фонем русского языка требует одиннадцати признаков, система индийского языка пали — десяти, а для описания языков Полинезии, где согласных звуков очень мало, достаточно пяти-семи различительных признаков... Но, быть может, в джунглях Новой Гвинеи или Амазонии отыщется язык, который включает все двенадцать признаков? Или даже потребует для своего описания еще одну, тринадцатую пару противопоставлений?

Кстати сказать, многие лингвисты полагают, что двенадцати признаков мало для того, чтобы описать языки, даже хорошо известные. И предлагают увеличить число различительных признаков до пятнадцати и даже более. Система того или иного языка начинает описываться с помощью так называемой объемной фонологии — связи между признаками, образующими каждую фонему, изображаются не на плоскости, а в пространстве трех измерений, стереометрически. Допускаются и более сложные, четырехмерные или пятимерные модели, которые могут описать более сложные соотношения фонем и их признаков.

Но главное не в том, чтобы отыскать новый различительный признак или построить точную модель фонем конкретного языка в двух, трех или нескольких измерениях. Мы начали наш рассказ о фонологии со сравнения. Фонемы — это атомы языка, из которых складываются молекулы слов и смыслов. С помощью различительных признаков, своеобразных элементарных частей, удалось построить подобие таблицы Менделеева, только не химической, а фонологической.

Как известно, ядерная физика до сих пор не имеет общей теории поля. Огромное число элементарных частиц, открытых в наше время, не поддается классифика-

ции и группировке. Некоторые физики предлагают понятие кварка, которое, по их мнению, является еще более базисным, элементарным, чем классические элементарные частицы...

Быть может, не за горами то время, когда и в фонологии появятся свои фонологические кварки? И найдены они будут уже не акустиками, изучающими звуковой поток человеческой речи, и не физиологами, исследующими аппарат порождения этой речи, а вероятно, учеными, которые проникают в глубины нашего мозга.

Известный нейрохирург У. Пенфилд провел интереснейшие эксперименты, пытаясь отыскать в человеческом мозгу программы владения речью. Английские *t* и *d* различаются по признаку глухость — звонкость, а фонемы *h* и *k* — по признаку прерывность-непрерывность. Во время операций, при электрическом возбуждении коры головного мозга, эти противопоставления были сняты и вместе с ними исчезли и смысловые противопоставления. Вместо *бабочка* пациент Пенфилда говорил *моль*, вместо *ножницы* — *резаки*, вместо *постель* — *сено*... То есть не только фонетические, но и смысловые различительные признаки снимались.

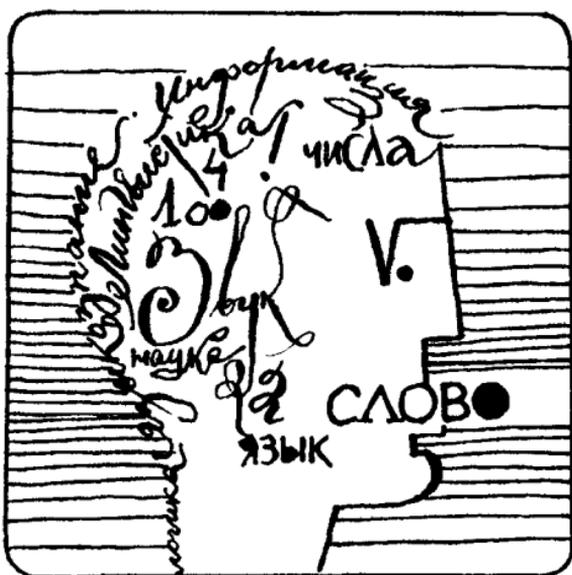
Это уводит нас к самой сложной, но зато и самой увлекательной и перспективной проблеме современного языкознания: каким же образом наш мозг производит речь? Почему, несмотря на все тонкости, на всю противоречивость, неоднозначность, нелинейность языка мы все-таки понимаем друг друга? Короче говоря, каковы механизмы порождения нашего ежедневного и ежеминутного чуда — человеческой речи?

Рассказом о поисках этих механизмов мы и закончим нашу книгу.

«Черным ящиком» называют в кибернетике систему, устройство которой неизвестно. Мы получаем информацию лишь на входе и на выходе и на ее основании стремимся понять поведение системы. Именно таким «черным ящиком» является пока наш мозг, особенно когда мы имеем дело с языком. Проникнуть в его тайны пытается новая дисциплина, родившаяся на стыке психологии и языкознания — психолингвистика.

О путях познания языковой деятельности человека, о попытке проникнуть в тончайшие механизмы порождения речи расскажет очерк

## ЧЕРНЫЙ ЯЩИК



## СЕМЬ ЗАПОВЕДЕЙ

Каждый день мы общаемся друг с другом с помощью слов. Говорить для человека так же свойственно, как есть, пить, дышать, думать. Но редко кто задумывается над тем, что же происходит при нашем «говорении», таком, казалось бы, естественном, как процесс дыхания или, скажем, пищеварения. А когда люди начинают задумываться над этим, чем больше они узнают и размышляют, тем больше их удивление, недоумение, восхищение...

В самом деле: почему мы все-таки понимаем друг друга?

Вразумительного ответа на этот вопрос не дадут ни филолог, ни психолог. Все попытки научного объяснения чуда «говорения» — и соответственно чуда понимания нашей речи — как в прошлом, так и в наши дни являются лишь гипотезами. Причем объясняют эти гипотезы лишь какую-то часть нашего чуда, да и то не всегда полно и точно. Зато любой серьезный специалист — будь то психолог, языковед или представитель психолингвистики, родившейся на стыке науки о психике и науки о языке, — прекрасно понимает всю бездну сложностей, что встают перед ним, как только речь заходит о постижении механизмов речи.

Крупнейший специалист в области психолингвистики американский профессор Дж. Миллер удачно сформулировал это в виде семи «заповедей» — характеристик семи аспектов языка, которые «без труда поймет любой психолог, дерзнувший попробовать свои силы в психолингвистике» (добавив, что можно без труда удвоить и даже утроить их). Попробуем же изложить эти «заповеди» языком, понятным любому смертному, а не только психолингвисту.

«Заповедь» первая. Мы привыкли считать, что произносим одинаково одни и те же звуки, слова, фразы. «Повторите, пожалуйста», — просят вас, не расслышав или не поняв, что вы сказали, и вы повторяете. Но если записать звуковые колебания, то окажется, что при повторе они несколько иные, чем были в первый раз. И голосовые связки при повторении работали не точь-в-точь так, как в первый раз.

Более того, не существует двух отрезков речи, которые были бы одинаковы по своей акустике или физиологически. И тем не менее нас превосходно понимают, когда мы повторяем нашу речь. Почему? Ведь материальные ее носители — звуковые волны и производящий речь физиологический аппарат — неодинаковы при повторе!

Более того, в любом языке множество слов, звучащих одинаково, но имеющих разное значение (русские слова лук, ключ, брак, пол и т. д.). Произносим мы их одинаково, но понимаем — по-разному. Опять-таки почему мы понимаем правильно эти идентичные слова?

«Потому, что они отличаются по смыслу», — ответят нам, и ответ будет правильный.

Но тут и начинается «заповедь» вторая. Хорошо, а что такое смысл? Мы говорим: «слова имеют значение», а что это такое?

Казалось бы, все просто. Слово стол обозначает предмет мебели с четырьмя ножками, на котором едят, табурет — предмет мебели с четырьмя ножками, на котором сидят, и т. д. Но далеко не все слова такие простые ярлыки предметов и явлений окружающего мира. Это давно уже поняли и философы, и логики, и языковеды.

В связи с этим находится и третья «заповедь». Значение высказывания не есть сумма значений слов, в него входящих. Выражение круглый невежда не значит то же самое, что невежда круглый. Исследование значения изолированных слов не решит проблемы, так как слова оказывают взаимное влияние друг на друга. Вечное перо и павлинье перо — совершенно разные перья, как бы одинаково ни звучали и ни писались слова, обозначающие эти перья.

Иными словами, значения слов комбинируются не по простым арифметическим правилам, а по каким-то другим, пока что неизвестным. При соединении значений целое оказывается не просто больше суммы его частей, а порой качественно чем-то иным. Или во всяком случае не равно ему (кстати, это одна из особенностей человеческого восприятия, работы нашего мозга по опознаванию образов и т. д.).

Это становится наиболее ясным, когда мы обращаемся к предложению, вершине языковой иерархии (фонема — морфема — слово — фраза). Отдельное слово обладает несколькими значениями. Но в контексте мы выбираем одно из них. Какое? Это зависит уже не от самого слова, а того предложения, в котором оно стоит. И даже от места этого слова в предложении.

Во многих языках мира — в тибетском, китайском и полинезийских — нельзя опознать существительное, прилагательное или глагол, пока не определим, чем является слово в предложении — подлежащим, определением или сказуемым. Да и в русском языке для понимания смысла слов важно знать их группировку. Иными словами, значение слов определяется во многом синтаксисом.

Это — четвертая «заповедь». А с ней связана пятая, гласящая, что число различных предложений практически бесконечно. Стало быть, бесконечно и число значе-

ний, которые могут быть выражены. Как же мы говорим и понимаем друг друга?

По всей видимости, существует некое порождающее устройство, что позволяет из ограниченного числа элементов-слов по определенным правилам грамматики строить бесконечное множество предложений. Причем порождение фраз подчиняется принципам, отличным от жесткой программы автомата или врожденной сигнальной системы, которая характерна для животных. Ведь в разговорной речи мы очень часто нарушаем правила языка: неточно произносим фонемы, комкаем слова, опускаем связки и другие грамматические элементы в предложении... И тем не менее понимаем друг друга.

Все попытки сформулировать эти правила в виде законов языка до сих пор успеха не имели.

Более того, как подчеркивает Дж. Миллер (и это шестая его «заповедь»), описание правил, которыми владеют все знающие язык, отличается от описания психологических механизмов, действующих при использовании этих правил. Описание языка, его структуры — не есть описание носителя языка, черного ящика, который воспринимает, хранит, перерабатывает и передает информацию. И, как замечает Миллер, «пределы кратковременной памяти самым серьезным образом ограничивают нашу возможность следовать нашим же собственным правилам». А ведь при разговоре мы пользуемся именно этой кратковременной, или оперативной памятью, ибо передача информации идет быстро, почти на пределе пропускной способности обработки информации мозгом.

И, наконец, последняя седьмая «заповедь» Все попытки обучить животных говорить по-человечески были неудачны. По всей видимости, эта задача неразрешима (хотя зоопсихологи с успехом обучают шимпанзе «языку жестов», «языку знаков», но не звуковому языку людей!). Как говорят специалисты, чтобы усвоить человеческий язык и чтобы пользоваться им, нужно быть человеком.

«Как это бывает в науке, можно считать, что мы наполовину победили, если начали правильно формулировать вопросы,— заключает свои «заповеди» Дж. Миллер. — Однако интереснее всего создать некоторую позитивную программу исследований, если мы действительно хотим получить реалистическое представление о

том, что такое язык... Если гипотетические построения, которые необходимы для исследования, кажутся вам чересчур сложными и нестрогими, чересчур невероятными и надуманными, вам лучше отказаться от мысли исследовать язык. Потому что язык сложен, произволен и осмыслен и никакие теории относительно языка не сделают его иным».

Итак, все, что я хотел подчеркнуть своими «заповедями», говорит Дж. Миллер, следующее: язык невероятно сложен.

И все-таки ученые пробуют решить загадку этого невероятно сложного языка.

### «ЯЗЫК В КОЛЫБЕЛИ»

«Вклад Энтони в лингвистическую теорию» — так называется предисловие, написанное к книге Руфи Вейер профессором Гарвардского университета Романом Якобсоном. Книга называется «Язык в колыбели». А Энтони — это двухлетний мальчик, чья речь в кроватке в течение нескольких месяцев записывалась на магнитофон его мамой Руфью Вейер. Обработка этих записей легла в основу книги, которую штудируют лингвисты, психологи и прежде всего психолингвисты. Ибо становление языка — это один из самых важных и надежных «ключей», с помощью которого можно проникнуть в загадку нашего обыденного чуда — человеческой речи.

Выше мы уже обсуждали с вами вопрос о том, как ребенок овладевает речью. Логика повествования о языке как системе подводит нас снова к этой теме.

Итак, человек не рождается с умением говорить, и без помощи людей ребенок языком не овладеет. Об этом написано даже в учебнике «Русского языка» для четвероклассников. Но вместе с тем у человека есть врожденная способность говорить.

С момента своего рождения начинает издавать звуки младенец. Крики беспокойства и звуки удовлетворения — так отвечает новорожденный на все события, происходящие вокруг. Конечно, с человеческой речью они не имеют ничего общего, это чисто биологические, животные «крики нутра». Никто не обучал им ребенка, они встроены в нас природой.

Младенец растет. Увеличивается его мир, растет и число разных звуков, которые издает малыш. Вначале

крики беспокойства и звуки удовольствия, если мерять их рамками нашей обычной речи,— это гласные звуки. Постепенно звуковой репертуар увеличивается: к крикам беспокойства добавляются *в, л, х, дж*, а к звукам удовольствия — согласные *г, к, р...*

Почему именно эти звуки? Лишь потому, что человеческое горло легче всего может издать их. Ребенок не произносит, а только издает звуки, точно так же, как и любой другой детеныш животных, обладающий голосом.

И лишь позже начинается превращение инстинктивной, физиологической, «нутряной» речи в настоящую человеческую речь. Превращение происходит не само собой, а под влиянием других людей: отца, матери, окружающих близких. Детский лепет переходит в членораздельную речь. В два года ребенок знает около трехсот слов. В три года — около тысячи слов, в четыре — от полутора до двух тысяч слов.

Но, как говорил знаменитый немецкий языковед Вильгельм Гумбольдт, «усвоение детьми языка не есть приспособление слов, их складывание в памяти и оживление с помощью речи, но развитие языковой способности с возрастом и упражнением». Именно на развитие этой языковой способности и обращают в первую очередь свое внимание ученые.

Когда ребенок лепечет, никакие усилия не заставят его говорить по-настоящему. Он для этого еще не созрел. Если же, напротив, пропустить период «от двух до пяти» и не обучать ребенка языку, он рискует стать умственно отсталым и никогда не овладеть речью. Есть определенные оптимальные, критические периоды развития ребенка для усвоения языка: ни раньше, ни позже он его как следует не усвоит.

Как же происходит это усвоение? Обычно мы говорим, что ребенка учат говорить. А как, собственно говоря, учат? Самая обыкновенная мама понятия не имеет о всех этих фонемах, морфемах, синтагмах, порождающих моделях, с помощью которых описывают язык лингвисты. Не знает она и о моделях обучения, разрабатываемых психологами, о различных схемах подкрепления, звуковых реакциях и т. п. И все-таки русские, американские, бушменские, немецкие и ненецкие дети прекрасно овладевают родным языком! Как же это про-

исходит? В том-то и дело, что мы этого и по сей день не знаем...

Психологи считают, что есть два основных механизма обучения: подкрепление и подражание. Рассмотрим первое. Обучение ребенка языку существенно отличается от той системы подкреплений, которой пользуется ученый в лаборатории, экспериментируя с животными или взрослыми людьми, задавая им какой-либо специальный тест.

Родители поправляют ошибки ребенка, но делают это не все время. Сами родители и окружающие взрослые, не говоря уже о детях, порой говорят неправильно. «С первых же дней жизни ребенка на него обрушивается чудовищная смесь из грамматически правильных предложений, неправильных обрывков фраз, неполных высказываний с подразумеваемыми частями, вопросов и, наконец, сюсюканья, подражающего его собственным попыткам говорить,— пишут американские психологи П. Линдсей и Д. Норман в книге «Переработка информации у человека». — Из всего этого словесного винегрета ребенок должен как-то извлечь правила, управляющие передачей информации о понятиях и отношениях».

Мы не ведем постоянного контроля над «говорением» ребенка. Мы не оцениваем это «говорение» в терминах хорошо и плохо, а если и даем эти оценки, то не все время. Мы не имеем точного представления о том, чему же хотим научить ребенка, так как механизм «говoreния» и по сей день неизвестен даже ученым, не говоря уже о простых смертных — мамах, папах, бабушках и т. д.

И все же ребенок обучается говорить! Может быть, здесь срабатывает другой механизм обучения — подражание? Ведь у ребенка инстинкт подражания развит необычайно сильно. Может быть, слушая родителей и окружающих, ребенок начинает подражать их речи и таким образом овладевает ею?

Однако, как заявляют психологи, вся последовательность процесса усвоения языка свидетельствует как будто против этого. Нет, ребенок не может обучиться речи, лишь подражая тому, что он слышит. Первые слова, которые произносит ребенок, напоминают пародию на реальное звучание этих слов. Однако вместо того, чтобы

добиться точного воспроизведения того или иного слова, дети предпочитают калечить новые и новые слова.

«Ни на одной из стадий овладения языком, от первых однословных высказываний и вплоть до речи, подобной речи взрослых, модели детской речи, по-видимому, не являются подражанием услышанному. Напротив, самое поразительное в речи детей — это ее новизна, — пишут Линдсей и Норман. — Более того, ребенок с врожденными дефектами речевого аппарата научается великолепно понимать язык, хотя он не способен говорить и, следовательно, подражать услышанному».

В языке существует бесконечное множество фраз и выражений. Редкие из них слышит ребенок по нескольку раз, большинство предложений однократны. Ребенок, по всей видимости, обучается не с помощью подкреплений или подражания, а пользуясь самим языком, извлекая из речи систему правил, которая лежит в основе языка. И ухитряется извлекать, несмотря на то, что все попытки современной науки сформулировать эту систему, до сих пор, как уже не раз говорилось здесь, не дали удовлетворительных результатов.

Почему же это удается ребенку? «Трудно себе представить, чтобы дети могли так быстро научиться языку от родителей, которые слабо себе представляют, что такое язык, если бы дети в результате долгой эволюции не были бы приучены выбирать аспекты языка, имеющие универсальную значимость, — констатирует Дж. Миллер. — Любое человеческое общество имеет язык, и у всех языков есть общие черты — черты, называемые «языковыми универсалиями», которые на самом деле по своему характеру являются долингвистическими... Это и есть, коротко говоря, важный биологический компонент, определяющий существующий характер человеческих языков».

И при всем при этом язык остается явлением прежде всего общественным, социальным. Генетически в нас встроены не сам язык, а способность к обучению человеческой речи.

## ПОРОЖДАЮЩИЕ ГРАММАТИКИ

«Изучение потока речи без гипотез о механизме его порождения не только малопродуктивно, но и не интересно», — говорит академик А. Н. Колмогоров, чьи

работы внесли большой вклад в моделирование языка и поэтического творчества.

В самом деле: можно без конца изучать методами статистики частоту употребления тех или иных слов, фонем, конструкций нашей речи. Можно строить самые разнообразные структурные модели, описывающие грамматику или синтаксис того или иного языка. Но ведь число фраз, не говоря уже о текстах, практически бесконечно... Нельзя ли попробовать описать язык не как статичную структуру, а как некое динамическое устройство, которое порождает любое число предложений из ограниченного числа элементов—слов и правил их сочетания? Или любое нужное число слов из элементов более низкого уровня — морфем и фонем...

Американский лингвист Ноам Хомский в течение нескольких лет вел плановую академическую работу по структурному описанию языка. Но чем дальше продвигалась эта работа, тем ясней ему становилось, что этот путь ведет в тупик. Классической считалась следующая методика описания языка: записать на магнитофон как можно больше текстов у носителей того или иного языка (например, исчезающих на наших глазах индейских наречий), затем провести своего рода дешифровку, выявить его структуру на основании этих текстов.

Такая методика оправдывает себя, если мы имеем дело с мертвыми языками, дошедшими до нас в виде писем или записей. Ведь к этим текстам невозможно добавить новые. Однако во всех живых языках, а их на планете, как мы уже знаем, несколько тысяч, число текстов практически неограниченно. Индеец или папуас, англичанин или русский в любое время выдаст нам сколько угодно текстов на своем родном языке. Хотя он может при этом и понятия не иметь ни о каких структурах, да и вообще о грамматике, семантике, фонетике.

Предположим, что в человеческом мозгу есть какое-то порождающее устройство, с помощью которого мы можем сказать обо всем, что видим, слышим, думаем, а благодаря другому устройству — понять то, что нам говорят. Нельзя ли применить формальный аппарат, которым располагает математическая логика, для описания не текстов, как это делает структурная лингвистика, а для описания самого процесса порождения речи?

Вопрос этот приходил в голову, конечно, не только одному Хомскому. Но именно Хомский смог, смело по-

рвав со сложившимися традициями, не только поставить вопрос, но и попытаться решить его практически. В математике существуют так называемые порождающие системы — формальные исчисления, с помощью которых, исходя из определенного набора аксиом, правил построения формул и вывода одних формул из других, можно строить бесконечное число правильно построенных высказываний... Нет ли тут аналогии с языком, с его бесконечным числом правильных предложений, состоящих из конечного числа элементов, слов и грамматических категорий?

Хомский принялся за работу. Итогом ее была рукопись в девятьсот страниц, излагающая принципиально новый подход к человеческому языку... которая была решительно отвергнута многими солидными издательствами. Однако в 1957 году голландское издательство «Мутон» выпустило краткий конспект этой рукописи под заглавием «Синтаксические структуры». Книга эта послужила отправной точкой для того явления, что ныне в лингвистике называют «хомскианской революцией». Книгу перевели на многие языки мира, в том числе на русский, на английском языке она выдержала почти десятков изданий. А вслед за нею появились новые работы самого Хомского и его учеников. Создание порождающих грамматик — одна из самых интересных и перспективных областей современной лингвистики. Более того, эта область дала толчок к развитию новых разделов математики, так называемой теории формальных грамматик. Вклад Хомского в науку сравнивают иногда с расшифровкой генетического кода, а самого автора теории порождающих грамматик некоторые увлеченные его последователи включают в список великих творцов нашего двадцатого столетия, состоящий уже из тысячи имен!

Подробный рассказ о порождающих грамматиках потребовал бы знакомства с математической логикой, структурной лингвистикой и многими другими областями знания. Но мы остановимся лишь на связи теории Хомского с непосредственной темой нашего очерка — с работой человеческого мозга, порождающего и воспринимающего речь.

«Подобно тому, как мы вывели необходимость существования трехмерного пространства, скрытого за изображением на сетчатке, мы должны обнаружить синтаксические структуры, лежащие в основе линейной

цепочки звуков, образующих предложение. Исследователь пространственного восприятия должен хорошо разбираться в проективной геометрии и столь же хорошо должен разбираться психолингвист в грамматике», — говорит Дж. Миллер. И добавим мы от себя: работы последних лет показали, что не меньших знаний требуется от психолингвиста в области семантики, смысла слов и фраз. На первых порах казалось, что между синтаксисом, который описывается в терминах теории порождающих грамматик, и семантикой лежит непреодолимая пропасть. Однако в наши дни делаются попытки описывать значение столь же строгими методами (вспомните очерк «В поисках значения»). Для семантики строятся свои порождающие модели, которые стыкуются с порождающими моделями грамматики. Ибо, по признанию последователей Хомского, понимание фразы столь же сильно зависит от контекста, в который она включена, сколько от ее синтаксической формы.

Сам Хомский считает свою теорию лишь начальным подходом к проблеме «понимания всей богатой области языкового опыта». Порождающие модели грамматики или значения пользуются строгой и однозначной терминологией математической логики. Формулировки их понятны и нам, людям, и электронному мозгу машины, ибо это алгоритмы, в которых описываются правила языка. А язык, как вы сами прекрасно знаете, гораздо гибче, он имеет колоссальную свободу выбора. «Операции, совершаемые механизмом речи в процессе ее порождения, не могут носить во всех случаях абсолютно «жесткий», автоматический характер. Они необходимо перемежаются с вероятностными операциями, в ходе которых совершается отбор одного из возможных путей продолжения процесса», — пишет советский лингвист С. Д. Кацнельсон. Помимо алгоритмов, мозг руководствуется еще и вероятностями! В первом очерке мы говорили о том, что современная лингвистика и теория знаков различают язык и речь, систему и тексты, этой системой порождаемые. Структура текстов описывается с помощью статистики, за многообразием фактов-чисел ученые пытаются увидеть формулы, механизмы языка. И по мере того как растет наше знание этих механизмов, мы все больше убеждаемся в том, что и сами механизмы порождения речи подчиняются не столько однозначным «алгебраическим», сколько вероятностным правилам и законам.

Человеческий мозг очень часто сравнивают с вычислительной машиной. Но машина эта действует не по строго заданной программе жестких алгоритмов. Ведь мы живем в постоянно меняющейся среде, никто не в состоянии учесть все случаи, ситуации, изменения, которые могут произойти в окружающем нас мире. Наше поведение, управляемое мозгом, носит вероятностный характер. Мы строим в уме свособразную модель будущего, опираясь на наш прошлый опыт и на те пробы и прощупывания, которые условно обозначаются как ориентировочные реакции.

Порой мы делаем это сознательно: вычисляем вероятность того или иного события и принимаем соответствующее решение. Чаще же всего мы делаем это подсознательно, «прикидывая» вероятности, хотя и не выраженные в точных числах, избирая стратегию поведения в соответствии с этими вероятностями, и делаем определенные «ходы», хотя не имеем ни малейшего представления о математической теории игр, моделирующей поведение. Как выразился один из корифеев советской физиологии, профессор Н. А. Бернштейн, в любой фазе, намечая пути дальнейших действий, «мозг в состоянии лишь наметить для предстоящего момента своего рода таблицу вероятностей возможных исходов».

Скорее всего, как вероятностная машина работает наш мозг и тогда, когда он воспринимает человеческую речь. Ведь восприятие ее идет на фоне самых различных «помех»: мы можем неточно произнести тот или иной звук, «проглотить» слог, опустить связку, употребить многозначное слово, упростить синтаксическую конструкцию в устной речи. Все эти помехи при декодировании, восприятии речи мозгом ликвидируются, причем очень быстро.

Математическая лингвистика находит, что язык имеет свои четкие статистические параметры, частотные характеристики того или иного стиля, употребления падежа, слова, грамматической конструкции... Может быть, подобного рода вероятности существуют не только объективно, в текстах, но и субъективно? То есть своеобразная и уникальная вычислительная машина — мозг — имеет свою собственную вероятностную модель

языка? И согласно этой модели происходит восприятие и порождение речи?

Проверкой этой гипотезы занимались многие ученые в нашей стране и за рубежом. И вот какие интересные результаты были получены ими...

Возьмем русский язык. С помощью вычислительной техники составлен частотный словарь русского языка, на основании которого мы можем судить об употребительности того или иного слова. Из этого словаря была взята выборка в сто слов, имеющих разную частоту употребления. Затем тридцать человек разного возраста, профессии и уровня образования рассортировали эти слова по семи категориям — в зависимости от того, как часто в их языковой практике они встречаются (никогда — очень редко — скорее редко, чем часто — не очень редко, но и не часто — скорее часто, чем редко — весьма часто — на каждом шагу).

Обработка результатов эксперимента показала, что в основном частоты субъективные, хранящиеся в мозгу носителей языка (в данном случае русского, но подобного же рода опыты были проведены и по другим языкам), совпадают с частотами объективными, полученными в результате статистической обработки текстов на ЭВМ.

Правда, были некоторые расхождения. Касались они прежде всего общеупотребительных, бытовых слов, таких, как туфли, зонтик, пакет и т. п. По данным частотных словарей, слова эти очень редкие. А по субъективным оценкам — скорее частые, чем редкие... В чем тут дело? Очевидно, в том, что при оценке слов этого типа участники эксперимента опирались не на частоту самого слова, а на частоту встречаемости предмета, этим словом обозначаемого. Предметы вроде туфель, зонтика и т. п. в нашем быту встречаются очень часто, а слова, их обозначающие, — редко.

Например, проанализировав частотный словарь французского языка, составленный на базе полутора миллионов слов, ученые обнаружили, что там не оказалось таких обиходных слов, как зубы или ножницы. В частотном словаре русского языка, составленном на базе миллиона слов, нет таких житейских слов, как расческа. Анализ устной речи свидетельствует, что и в ней слова типа туфли или ножницы не являются частыми. А дальнейшие эксперименты психолингвистов показали,

что расхождение между объективными и субъективными частотами приходится, как правило, именно на эти обиходные слова — мы склонны завышать их частоту, ибо предметы употребляются в быту очень часто (чего нельзя сказать о самих словах!).

Кто же более прав — объективные частотные словари или наша языковая интуиция, считающая обиходные слова частыми, а не редкими? По данным подсчетов по текстам в устной или письменной речи, обиходные слова всегда попадают в число редких. Однако в нашем представлении они являются частыми. И, как отмечает в своей монографии «Вероятность элементов текста и речевое поведение» Р. М. Фрумкина, субъективные оценки частоты, «по-видимому, лучше отражают вероятностную организацию словаря в речевых механизмах испытуемых, чем данные подсчетов по текстам». И в этом смысле являются более сильным прогнозирующим фактором, чем «объективная» оценка частоты, полученная с помощью статистической обработки текстов на ЭВМ.

### В ДЕБРЯХ АССОЦИАЦИИ

Исследования показывают, что подобного рода вероятностные оценки имеются у нас не только для слов, но и для слогов, сочетаний звуков (или букв) и даже отдельных звуков или букв. «Носители языка обладают способностью оценивать частоты элементов текста меньших, чем слово — букв, двухбуквенных и трехбуквенных сочетаний» — таков был вывод психолингвистов, проводивших опыты на разных испытуемых и на материале различных языков — русского, английского, французского и др.

Еще более интересные результаты были получены психолингвистикой, когда были проведены опыты не со словами или сочетаниями букв, а в иной плоскости — смысловой. Мы уже рассказывали об измерении значений. Но помимо субъективной, оценочной характеристики слов, слова в нашем сознании связываются с другими словами по ассоциации. Вот характер этих ассоциаций и попробовали выявить психолингвисты. И с первых же шагов столкнулись со множеством трудностей, неизвестных до того ни психологам, ни лингвистам. Причем трудности эти возникли, как говорится, с

порога: каким образом классифицировать ассоциации, которые человек дает на то или иное слово?

Тут и синонимы (на слово *боль* дается ответ *страдание*) и антонимы (*мокрый* — *сухой*) и подчиненные слова (*фрукт* — *яблоко*), и соподчиненные (*яблоко* — *персик*), и подчиняющие (*яблоко* — *фрукт*), и дополняющие (*вперед* — *продвижение*), и грамматические изменения слов (*иду* — *шел*) и «эгоцентризмы» (на слово *успех* дается ответ: *я должен* и т. п.). Причем большое число слов нельзя отнести к какой-либо определенной категории (например, в ответ на *приказ* дается: *и весь сказ* — что это? эгоцентризм? дополняющие слова? или ответ в рифму, по созвучию?).

Однако, несмотря на все трудности «объективного исследования тех смысловых связей, которые реально возбуждаются у человека тем или иным словом» (так называют иногда подобного рода работы), ученым удалось установить бессознательные связи между словами, которые существуют в голове носителя русского, английского и любого другого языка мира. Связи эти должны отражать и связи понятий, этими словами выражаемых.

Возьмем классический пример. Заставьте испытуемого вести счет вслух и не подряд, а двойками или тройками, то есть 2—4—6—8—10 и т. д. или 1—3—6—9—12 и т. д. Затем попросите его называть слова-ассоциации в ответ на слова *птица, поэт, дерево*. Почти наверняка ответы будут звучать так: *курица, Пушкин, яблоко*. Настолько в нашем мозгу связаны между собой эти понятия! (Кстати, в экспериментах американских психолингвистов слово *tree*, дерево, оказалось теснейшим образом связано со словом *apple*, яблоко.)

Интересные результаты получил американский ученый Дж. Диз. Сначала он взял несколько десятков слов, выступающих как самые частые ассоциации на слово *butterfly*, то есть бабочка. Слова эти, в свою очередь, послужили исходными для новых ассоциаций. Затем полученные результаты были обработаны математически, с помощью факториального анализа. И они показали несколько различных факторов.

Первый связан с обозначениями живых существ: *моль, птица, муха, пчела, жук, насекомые* и т. п. Второй фактор — в словах, противоположных первому, в обозначениях неживых объектов вроде *сад, цветок, солнечный свет, природа, лето, синий, желтый* и др. Третий

фактор позволяет разделить живые существа на положительные и отрицательные: *пчела — хорошая, моль — плохая*... Четвертый фактор делит на две группы неживые объекты. В одной группе оказались слова *синий, небо, желтый*; в другой — *лето, солнечный свет, цветок*.

Опыты показали, что связи по ассоциации имеют не только смыслы отдельных слов, но и категории грамматики. В ответ на существительное чаще всего называется также существительное. Хотя и здесь все не так просто. Например, в ответ на существительное *небо* восемь из десяти человек дают ответ прилагательным *голубое*. А в то же время *земля* вызывает ассоциации со словами *небо, вода* и другими существительными. Прилагательные же вроде *черный, твердый* называются крайне редко.

Характер ассоциаций у взрослых и детей различен. Взрослые почти в десять раз чаще называют ассоциативные слова по контрасту, чем дети. Мы уже рассказывали в очерке, посвященном поискам значения, об эксперименте, проведенном в нашей стране. Речь шла о числовых индексах, с помощью которых можно выразить степень связи слов, противоположных по смыслу: *зрячий — слепой, черный — белый*... Эксперимент этот не закончился вычислением индексов.

«После эксперимента, результаты которого позволили вычислить индексы противопоставления к слову «лед», мы провели детальный опрос группы испытуемых в возрасте от 19 до 32 лет, как правило — с законченным средним образованием и частично — с незаконченным высшим, — пишет руководитель опытов А. А. Брудный. — Из общего числа участвовавших в эксперименте испытуемых была отобрана группа с таким расчетом, чтобы каждый контрастный ассоциат был представлен не менее чем двумя «авторами».

Что же заставляло людей называть то или иное слово как ответ — противопоставление слову *лед*? Почти половина всех испытуемых отвечала словом *вода*. Почему? «Потому, что раньше льда — вода», — таков был один ответ. Второй, почти идентичный, звучал: «после льда — вода», третий — «потому, что вода жидкая, а лед твердый». Ответ *огонь* дала примерно пятая часть всех участников эксперимента. Объяснением его были слова: «Потому, что огонь — самое горячее, а лед — самое холодное». Около одной десятой опрошенных

ответило словом *пламя*. И этот ответ получил такую же мотивировку, что и ответ *огонь*.

Казалось бы, ответ *пламень*, столь же частый, как и *пламя*, должен объясняться так же. Однако объяснения были иными: «так у Пушкина» и «лед и пламень — крылатые слова». Тут, видимо, действовали уже не просто словесные или понятийные, а чисто литературные ассоциации.

Ответ *кипяток* получил мотивировку: «он горячий и жидкий», ответ *пар* — «это самая легкая вода, а лед — наиболее тяжелая», ответ *снег* — «это противоположность зимних состояний воды» и «он мягкий». Ответ *песок* объяснялся словами «представил пустыни и льдины» и «он сыпуч, а лед — ломок». Ответ *земля* — «где земля, льда нет, лед, где вода» и «летом: земля — вода, а зимой: земля — лед». Ответ — *жар*: «жар противоположен льду, потому что лед — концентрат холода». Наконец ответ *таянье* обосновывался тем, что «это гибель льда».

Как видите, в ответах-мотивировках отразились самые различные свойства льда: его вещественность (твердость, ломкость, холодность), его бытие (лед — реализованное свойство воды замерзать и таять), его символика (противопоставление качеству — пламень; лед как сгущенный холод; лед как нечто неподвижное, сменяющее зимой летнюю воду, бегущую и теплую, подобно тому, как смерть сменяет жизнь)... Вот каким одновременно и ориентированным и неоднозначным оказывается значение слов, которые мы употребляем, казалось бы, автоматически и запросто!

А ведь ассоциации на слово *лед* давались, так сказать, направленно — требовалось назвать слово, противоположное по смыслу, и только. Можно представить, какими же сложными и многомерными связями оперирует наш мозг, когда подбирает нужное слово или выражение. Причем характер ассоциаций связан не только со структурой языка, но и определенными нормами данной культуры, а также и личными интересами конкретного человека. Примерно восемьдесят поляков из тысячи в ответ на слово *музыка* скажут: Шопен. Американцы же дают такой ответ лишь в двух случаях из тысячи, зато чаще, чем поляки, назовут Бетховена. Почти треть всех испытуемых в США на слово *статуя* давало ответ: Свободы. Носители английского языка в

любой другой стране, разумеется, столь частого ответа не дадут.

Ассоциации могут быть связаны и с событиями текущего дня. Порой то или иное слово, название нашумевшего романа или кинофильма включается в нашем сознании в сферу значений слов. Возьмем словарь ассоциаций, выпущенный Дж. Джироу и Г. Поллио в 1965 году. Многие американцы на слово *страх* отвечали словами *война* и *Вьетнам*. На слово *a controversy*, означающее спор, дискуссия, полемика, давался ответ — *расовый, сегрегация* и даже «Лолита» — по наименованию скандально известного романа Владимира Набокова!

### АВТОМАТ ПАРАЛЛЕЛЬНОГО ДЕЙСТВИЯ

Ассоциации, как вы, вероятно, сами убедились, могут быть сложны и прихотливы. Однако — и это, видимо, вы также заметили, — они вращаются в определенной сфере, в каком-то смысловом «поле». Психологи выделили больше двух десятков типов словесных ассоциаций, которые затем сгруппировали в более общие категории: сходство (темный — черный, гора — холм); смещение (стол — стул, дом — сарай, то есть отнесение к определенной категории — мебели, построек); контраст (темный — светлый, большой — маленький — словом, ответы-противопоставления); расширение (стол — мебель, дом — постройка) и сужение (фрукт — яблоко, мебель — стул). Еще один тип ассоциаций — перенос значения. Например, причина — следствие: болезнь — смерть, преступление — наказание; часть — целое: стол — ножка, солдат — армия; предмет — материал: стол — дерево, лампа — стекло; действие — объект; вбивать — гвозди, ставить — вопросы; действитель — объект: паук — паутина, ученый — наука...

Помимо этих, связанных определенными смысловыми отношениями, ассоциаций существует еще один тип. Слова называются не по своему значению, а по своему звучанию, по созвучию, рифмуемости и т. п. Причем это делают не профессиональные поэты, а обыкновенные люди. И вывод, который отсюда следует, прост: в нашем мозгу существуют не только смысловые, но и звуковые связи слов. Именно эти связи использует поэт, когда пишет стихи. И, вероятно, именно они помогают слуша-

телям поэзии воспринимать поэтическое творчество, сопереживать не только мыслям и чувствам поэта, но и ритму, рифме, звуковой инструментровке стиха.

До сих пор речь шла, так сказать, об обыденном чуде — нашей прозаической обиходной речи. Но ведь существует еще и «чудо в квадрате» — поэтическая речь. Давайте-ка обратимся именно к ней, ибо тут, несмотря на всю сложность и многомерность поэзии, наглядно видны свойства языка вообще, которые в обычной практике мы не замечаем.

Знаменитый русский сатирик Салтыков-Щедрин, ведя полемику с эстетствующими поэтами, как-то едко спросил: «Зачем ходить по канату, приседая на каждом четвертом шагу?» Однако, как показали исследования стиха методами теории информации, несмотря на все «приседания», поэт ухитряется дать гораздо больше информации, чем обыкновенные смертные, пользуясь тем же языком и теми же звуками, буквами, словами. Свободный синтаксис, смелость образов, стремительность изложения, гибкость языка, допускающая различные варианты передачи сообщения, — все это позволяет поэту передать не только смысл, но и вдобавок волевою и эмоциональную информацию (в технической кибернетике, замечает академик А. Н. Колмогоров, первая находит свой аналог в управляющей информации; эмоциональная же информация в технической кибернетике своего аналога не имеет).

Специалисты уже давно спорят о различиях между прозой и стихом. К единому соглашению они так и не пришли. Вполне возможно, что вообще нельзя провести такую четкую разделительную черту: вот от сих начинается поэзия, а вот от сих — проза. Но в самых обычных и типичных случаях различие между стихами и прозой состоит в том, что когда мы воспринимаем стих, то явно ощущаем закономерности, которые имеются в воспринимаемой нами речи. Мы не только ощущаем их пассивно, мы переживаем их открытие, мы «соучаствуем». Наше подсознание вовлекается в активную работу: оно прослеживает и проверяет эти закономерности, а как только наталкивается на перебой, то сразу же сигнализирует сознанию об этом. Причем интуитивное восприятие стиха необычайно тонко. Фраза «брат упрямил награду дать» возможна в четырехстопном классическом ямбе, а вот фраза «брату просил награду

дать» — нет. Наше подсознание, воспринимающее стих, знает, оказывается, правила грамматики!

А ведь ритм — это только нижний, первый этаж поэзии. Далее следуют более высокие этажи: инструментовка, наполнение ритмической «решетки» звуками. Но, как известно, поэт пишет не звуками, а словами, и выбор слов — следующий этаж, далеко еще не самый верхний. Слова сочетаются друг с другом, образуют предложения, а те служат материалом для создания поэтических образов, сюжета, эмоциональной окраски всего произведения, выражения его общего настроения, поэтической мысли. Наконец, самый верхний этаж — это осознание поэтом окружающей действительности, или, говоря языком кибернетики и теории знаков, создание «модели мира», в которой поэт выразил свое отношение к социальным, психологическим, историческим, «биографическим» и многим другим факторам.

Каким образом ухитряется человек, пишущий стихи, оперировать одновременно и смыслом, и звуком, и ритмом, и ассоциациями, и многим, многим другим? Очевидно, здесь идет какая-то комплексная, неведомая и по сей день работа, еще более сложная и удивительная, чем та, которая происходит при нашем обычном прозаическом «говорении».

### ВНУТРИ «ЧЕРНОГО ЯЩИКА»

Как работает наш мозг, этот кибернетический «черный ящик», во время порождения и восприятия речи, мы до сих пор не знаем. Однако в ряде случаев ученые имеют возможность хоть краешком глаза заглянуть внутрь этого «черного ящика». Это случается тогда, когда у человека поврежден мозг и одновременно происходит расстройство речевых функций.

Разумеется, учеными движет не любопытство, а прежде всего желание помочь больному — человеку, который может мыслить и чувствовать, как все люди, но не в состоянии говорить или, наоборот, воспринимать речь окружающих. Болезнь эта называется афазией, и медицина знает множество вариантов афазии, в зависимости от того, какой раздел головного мозга поврежден: ложный, теменной, височный. В прямой связи с поврежденным бывают и расстройства речи: больной может говорить, но не воспринимает речь других; больной

понимает речь, но сам не может говорить правильно; больной говорит правильно, но не в состоянии координировать смыслы отдельных слов, сочетать их в предложении.

Афазии самых различных видов поддаются лечению. Причем в этом существенную помощь врачам начинают оказывать лингвисты. В последнее время афазии заинтересовали и кибернетиков, пытающихся научить «электронный мозг» языковым операциям. Уже первые опыты машинного перевода показали, что ошибки, которые делали ЭВМ, сопоставимы с теми, что делают больные афазией.

Сам процесс лечения афазий, обучения человека языковым программам, когда-то записанным в мозге, а затем нарушенным, заставил обратиться к сокровенным глубинам нашего мышления, базирующегося на языке. К чести нашей науки надо сказать, что ведущая роль здесь принадлежит советским ученым во главе с Александром Романовичем Лурия (недавняя смерть А. Р. Лурия не оборвала исследований, ибо после него осталась признанная во всем мире школа). Врачи, лингвисты и специалисты по психолингвистике нашли общее поле деятельности. Особое внимание психолингвистов привлекают афазии, связанные с нарушениями смысла.

При сенсорной (височной) афазии сохраняется способность воспринимать общие, абстрактные смысловые отношения. Однако ближайшие значения, оттенки смысла не различаются. Больной говорит «работать с пожаром» вместо «работать с огоньком». На вопрос, что такое тайга, он дает ответ: «что-то лесное... лесное...» Футбол для него «что-то физкультурное, а что?»

У больных моторной (лобной) афазией разрушается система значений, хотя конкретные значения остаются. Больной может опознать собаку, курицу, кошку, крысу. Он не в состоянии обобщить, что все они — животные.

Наконец, у больных семантической (теменной) афазией, говоря словами профессора А. Р. Лурия, «непосредственная предметная соотнесенность слова остается сохранной, вся же кроющаяся за словом система связей и отношений оказывается глубоко нарушенной». Значение слова лишается всего комплекса своих связей, оно распадается и, стало быть, легко забывается. И тогда, свидетельствует Лурия, в поисках нужного слова больной начинает испытывать те же затруднения, какие

испытывает нормальный человек при попытках вспомнить лишенную системы логических связей фамилию, часто заменяя искомое слово «случайными парафазиями», набором бессмысленных сочетаний звуков, дающих что-то «похожее» на искомое слово.

Под руководством Лурья были проведены эксперименты и на здоровых людях, связанные с поиском «хранилищ значения» в нашем мозгу. Современная аппаратура позволяет регистрировать сужение и расширение сосудов головного мозга. Сначала у нормального взрослого человека вырабатывался условный рефлекс на определенное слово. Например, на слово *кошка*. Затем давались другие слова в связи с ним и объективно фиксировались реакции, которые они вызывали.

Оказалось, что слова типа *облако, карандаш, стекло* и т. п. никаких реакций не вызывали. Точно так же, как и созвучные *кошке* слова вроде *крошка, крошка, окошко*.

Но как только назывались слова *котенок, мышь, собака, животное*, тотчас же фиксировалась реакция на них. И, что самое замечательное, — смысловые связи внутри нашего «черного ящика» оказывались порой отличными от тех, которые по логике относятся к той или иной категории (например, слово *арфа* не причислялось никем из испытуемых к струнным инструментам!).

Какие же механизмы действуют в нашем мозгу, когда мы подбираем слова для выражения той или иной мысли? Да и слова ли мы подбираем? Как считает ведущий советский специалист в области психолингвистики А. А. Леонтьев, слово в нашей памяти записано в форме поиска этого слова. Оперируя соответствующими признаками, мы тем самым уже «считываем запись» в лексиконе. И едва ли имеет смысл искать где-то в нервных клетках энграмму звуковой формы слова, какой-либо «отпечаток» с привешенным к нему «ярлычком». Ибо слово есть его поиск!

Ту же мысль зарубежные коллеги Леонтьева М. Анисфельд и М. Кнаппу выражают так: «Слова не хранятся в памяти как слова, но как комплексы признаков. Когда слова используются, они не репродуцируются памятью, а скорее реконструируются из составляющих эти слова признаков». Признаки же эти включают смысловые, грамматические, звуковые — фонологические и у грамотных вдобавок орфографические моменты.

Однако, если это действительно так, то слово оказывается уже не вещью, не ярлыком, а процессом; или, как отмечает Леонтьев, «если брать его более широко, как психологический эквивалент «словарного значения» — и вещь, и процесс, но никак уже не только вещь». А это означает, что мы должны коренным образом пересматривать сложившиеся представления о языке и знаковых системах.

### МИР, ЯЗЫК И МЫ

«С точки зрения сходства с реальным языковым поведением говорящего ни одну из известных до сих пор моделей нельзя признать удовлетворительной, — пишет известный американский психолингвист Д. Уорт. — Можно ли найти такую модель, которая соотносилась бы с действительным поведением реально говорящего (т. е., если можно так выразиться, «психосоциологическую» модель речи и языка)? Нам кажется, что да. Такая модель имела бы форму телевизионного экрана, с которым связаны два механизма, из которых один способен развертывать на экране разные изображения, а другой способен читать и различать эти изображения, передавать результаты чтения в «черный ящик», содержащий грамматические правила этого языка; «черный ящик» обрабатывает полученную от читателя — разлагательного механизма — информацию и передает результаты своей обработки первому, развертывающему механизму, который изображает на экране новую «картину»; этот циклический процесс продолжается (с электронной быстротой) до тех пор, пока «черный ящик» перестанет прибавлять новую информацию; весь аппарат тогда находится в состоянии стабильности, и картинка (т. е. предложение) «снимается» (т. е. говорящий произносит свое предложение)».

Уорт сформулировал свою модель в терминах кибернетики и электроники. Но, как известно, скорость протекания процессов в нервных волокнах в тысячи раз медленней, чем скорость электронов. И тем не менее человеческий мозг ухитряется опережать электронную машину даже в ее наиболее эффективной деятельности — счете. Соревнования между состоящими из плоти и крови чудо-счетчиками, людьми, и ЭВМ почти всегда кончаются победой человеческого мозга. Что же тут го-

ворить о языке — вы и сами, прочитав эту книгу, поняли, насколько совершеннее и сложнее человеческий язык самого сложного технического кода. А ведь модель Уорта не говорит еще о самом главном — о том, что «черный ящик» с помощью языка не просто передает информацию другому «черному ящику», но и познает, моделирует окружающий мир!.. Во многих современных работах употребляется выражение «модель черпака». Согласно этой модели смысл слов черпается изнутри сознания человека. Слово есть некий «черпак», единый для всех. Однако у разных людей далеко не одинаково содержимое, которое этим черпаком зачерпнуто.

Американский ученый Чарлз Лейярд в своей книге «Мысли о языке» приводит характерный пример. Вы собираетесь куда-то идти, а ваша спутница говорит: «Подождите минуточку».

Что такое *минуточка*? Астрономическая единица, равная шестидесяти секундам? Разумеется, не только это. В зависимости от обстановки — и характера вашей спутницы — это может означать, что действительно все готово и вот-вот вам составят компанию. Или, напротив, *минуточка* означает добрые полчаса и ждать надо долго. То есть смысл слова минуточка задается не только минутной стрелкой часов, но и ситуацией, и тем, кто говорит это слово.

Не является ли таким «черпаком» вообще наш человеческий язык? Чудодейственный черпак, с помощью которого мы извлекаем мысли из глубин нашего сознания и подсознания? Волшебный черпак, который служит обществу людей, связанных мириадами нитей друг с другом и окружающим их миром? Безотказный черпак, с помощью которого мы познаем мир, себя и, наконец, свой собственный инструмент познания — сам язык?

Ведь не будь языка, вряд ли автор смог бы изложить гипотезы и факты современной лингвистики, которой посвящена эта книга.

Первое издание книги «Звуков и знаков» ориентировалось в основном на знаки. Второе — в полном соответствии с идеями и поисками ученых всего мира, — на значение, на Его Величество Смысл, передача которого и является сутью нашей речи.

## 1878—1978 И ДАЛЕЕ... (Вместо послесловия)

Человеческий язык столь же древен, как и само человечество. В глубочайшую древность уходят попытки людей понять, что же такое их язык, как возникло это чудо. О происхождении языка и речи говорят мифы, записанные в пустынях Австралии и джунглях Амазонии, о чудесном даре богов — языке, именах, письменах — повествует мифология эллинов и шумеров, древних китайцев и египтян, индийцев и скандинавов.

Но не только мифами довольствовался человек. Сама его жизненная практика заставляла искать законы языка, своего и чужих. Как разговаривать с чужеземцем? Как записывать с помощью знаков звуки родной речи? Как правильно писать и говорить на родном языке и на языке иностранном? Как переводить тексты священных писаний, будь то Библия или стовосьмитомный буддийский канон, на языки язычников, дабы обратить их в свою веру?

Миссионеру и купцу, жрецу или учителю поневоле приходилось быть лингвистом. И уже несколько тысяч лет назад появились первые лингвистические сочинения. Это двуязычные словари, донесенные до нас глиняными книгами Двуречья. Это труды древнеиндийских грамматиков, во многом предвосхитившие идеи и методы современной структурной лингвистики. Это сочинения античных мыслителей, посвященные языку, смыслу имен и слов. Более девятисот лет назад замечательный средневековый ученый Махмуд Кашгарский написал свой «Диван турецких языков», материалом для которого послужили тюркские языки — турецкий, уйгурский и многие другие. В своей книге, ставшей известной европейцам спустя много веков после ее создания, Махмуд Каш-

гарский высказал идеи, которые лежат краеугольным камнем в фундаменте сравнительно-исторического языкознания, с которого, собственно говоря, и начинается подлинная наука о языке.

В конце XVIII столетия европейская наука открывает для себя священный язык Индии, санскрит. А вслед за тем ученые с изумлением обнаруживают, что язык этот поразительно похож на латынь и древнегреческий. И не только на эти мертвые классические языки, но и на персидский, исландский, русский, литовский и многие другие языки, распространенные в Европе и Азии. Оказывается, что все они имеют общего прапредка, все они ветви единого древа языков.

В течение XIX столетия Ф. Бопп, Р. Раск, А. Х. Востоков, Я. Гримм и многие другие ученые возводили стройное здание сравнительно-исторического языкознания. Лингвисты нашли соответствия между словами и звуками различных языков, родственных друг другу и образующих одну великую семью — индоевропейскую. Подобно тому, как палеонтологи по разрозненным костям восстанавливают облик вымерших животных, языковеды провели реконструкцию праязыка, предка всех индоевропейских языков — литовского или армянского, русского или персидского, английского или санскрита, исландского или цыганского...

Открытие древнего родства языков, разделенных тысячами километров, на которых говорят люди самых различных рас и культур, заставило ученых обратиться к истории этих людей и языков. Как получилось, что потомки викингов говорят на языках, родственных языкам горцев, живущих на Крыше мира — Памире? Почему языки древних эллинов и римлян находятся в родстве с языком жителей джунглей острова Шри Ланка, веддов, и по сей день живущих в каменном веке? Почему язык литовцев оказался поразительно близок к языку древних индийцев, отделенных от них тысячами километров пространства и десятками веков во времени?

Подобных вопросов возникало множество. Ответы же на них можно дать лишь с помощью языкознания, древней истории, этнографии, археологии — словом, только теснейший союз этих наук в состоянии пролить свет на многие загадки. И лингвистика заключила этот союз с историческими науками — союз, еще более окрепший в наши дни.

Индоевропейские языки были хорошо изучены и описаны. Найдены были формулы соответствий того или иного звука в различных языках и ветвях единого «древа языков», своею строгостью и точностью напоминающие формулы алгебры. Материалом для них были индоевропейские языки, древние и современные. Но вот ровно сто лет назад, в 1878 году, двадцатилетний дебютант в науке, молодой швейцарский лингвист Фердинанд де Соссюр выступает с заявлением, которое привело его маститых коллег в недоумение. По мнению молодого человека, в каком-то из индоевропейских языков должен быть некий звук, доселе неизвестный.

В каком именно? Соссюр такого языка назвать не может. Но тем не менее утверждает, что звук этот должен быть. К этому выводу его приводит анализ самой структуры звуков, системы языка. Иными словами, Соссюр предсказывает некий звук теоретически, подобно тому, как астрономы с помощью математики открывали новые планеты и кометы, как говорится, на кончике пера.

В ту пору Соссюру никто не поверил. И лишь полвека спустя, когда был найден ключ к языку таинственных хеттов, живших в Малой Азии несколько тысяч лет назад и говоривших на индоевропейском языке (сравните хеттское *небис* и русское *небо*, хеттское *дулуга* и русское *долгий* и т. д.), оказалось, что Соссюр был прав! В языке хеттов существовал звук, предсказанный им на основании анализа структуры языка. Это было ярким и наглядным доказательством того, что из науки описательной лингвистика может превратиться в науку гочную, способную не только описывать явления, но и предсказывать их! Вот почему год открытия Соссюра — 1878-й — справедливо считается годом рождения современного языкознания.

«Ныне нет лингвиста, который не был бы хоть чем-то ему обязан. Нет такой теории, которая не упоминала бы его имени, — пишет о Соссюре один из крупнейших современных лингвистов Эмиль Бенвенист, который, по словам профессора Ю. С. Степанова «наилучшим образом представляет современный этап науки о языке — лингвистику 70-х годов нашего века». — Лингвистика стала фундаментальной наукой среди наук о человеке и обществе, одной из самых активных как в теоретических изысканиях, так и в развитии метода. И эта об-

новленная лингвистика берет свое начало от Соссюра, именно в учении Соссюра она осознала себя как наука и обрела свое единство. Роль Соссюра как зачинателя признана всеми течениями, существующими в современной лингвистике, всеми школами, на которые она делится».

А направлений этих в наши дни, за истекшее столетие, появилось в лингвистике немало. Ибо современная наука о языке заключила союз не только с историческими дисциплинами, но и с математикой, психологией, техникой, естественными науками. Достаточно просто перечня, чтобы вы сами наглядно убедились в том, насколько многообразна современная лингвистика.

Теоретико-информационное изучение языка, анализ языка как удивительного и необычайного кода... Семиотическая лингвистика, рассматривающая язык как своеобразную систему знаков... Зоолингвистика, пытающаяся применить аппарат описания человеческого языка к описанию систем сигнализации, существующих в животном мире... Лингвопоэтика, рассматривающая поэзию и — шире — литературу как особым образом организованный язык... Стихотрофия, вносящая в изучение стиха строгие количественные меры... Фонология, открывшая «атомы» и «элементарные частицы» языка и ставшая в авангарде современной лингвистики... Экспериментальная фонетика, возникшая на стыке лингвистики, акустики и физиологии... Грамматология, наука о письме, преобразующем наш язык в условные знаки букв или иероглифов... Теория дешифровки, дисциплина, опирающаяся на грамматологию и лингвистику... Глоттохронология, стремящаяся определить темп изменения языка, найти своеобразные лингвистические часы... Лингвостатистика, измеряющая язык с помощью чисел во всех его проявлениях — от частоты фонем до устойчивых словосочетаний... Алгебраическая лингвистика, от использования традиционного аппарата математики перешедшая к созданию собственных «лингвистических исчислений»... Паралингвистика, изучающая явления, сопутствующие нашей речи — интонацию, жесты, мимику... Этнолингвистика, изучающая взаимосвязь языка, мышления и культуры... Неразрывно связанная с нею социоллингвистика, родившаяся на стыке языкознания и социологии... Геолингвистика, изучающая распространение языков на нашей планете, их «удельный вес» как

языков науки, дипломатии, культуры, международного общения... Психолингвистика, дисциплина, пограничная между психологией и языкознанием, в свою очередь, разделяющаяся на несколько дисциплин — теорию массовой коммуникации, ассоциативную лингвистику, методику «измерения значений» и т. д... Нейролингвистика, связанная с лингвистическим изучением расстройств речи... Патоллингвистика, позволяющая врачам-психиатрам на основании анализа речи больного давать точный диагноз психического заболевания... Инженерная лингвистика, развивающаяся в тесном содружестве лингвистов, математиков, программистов ЭВМ... Математическая лингвистика, создающая свой специальный аппарат вроде теории нечетких множеств и лингвистических переменных...

Наш перечень, далеко не полный, занял целую страницу. В очерках этой книги мы рассказали лишь о некоторых дисциплинах, самых актуальных, причем лишь об общей их направленности и проблематике. Подробный же рассказ только об одной из них, например инженерной лингвистике или фонологии, потребовал бы отдельной книги. Так же, как рассказ о социолингвистике, нейролингвистике, грамматологии и других дисциплинах, оставшихся неосвещенными на страницах «Звуков и знаков»... И все-таки, несмотря на все свое многообразие, современная лингвистика целостна и едина, ибо един ее объект, наш человеческий язык, наше повседневное чудо! Если же приглядеться внимательней, то окажется, что лингвистика XX столетия, несмотря на все новейшие идеи и технические достижения века, сохраняет кровную и неразрывную связь с традиционным языкознанием, с лучшими его представителями.

Вот несколько примеров такой преемственности. Еще в первой половине прошлого столетия замечательный ученый-энциклопедист Вильгельм Гумбольдт высказал мысль о взаимосвязи языка и мышления человека. «Тем же самым актом, посредством которого он (человек) из себя создает язык, он отдает себя в его власть; каждый язык описывает вокруг народа, которому он принадлежит, круг, из пределов которого можно выйти только в том случае, если вступишь в другой круг», — писал Гумбольдт.

Проходит почти столетие, и эта мысль на материале языков индейцев Америки развивается американским

лингвистом Бенджаменом Ли Уорфом. Гипотеза Уорфа и споры вокруг нее ведут к созданию этнолингвистики. А ее проблемы, как вы сами убедились, прочитав очерк «Модель мира», нельзя разрешить, если не обратиться к самому обществу, в котором живут носители языка, иными словами, без социоллингвистики.

Фердинанд де Соссюр по праву считается одним из основоположников современного языкознания. Идеи Соссюра о структуре языка в 30—40-х годах нашего столетия развиваются датским ученым Луи Ельмслевом, который показал, что между лингвистикой, математической логикой и теорией знаков, семиотикой, существует неразрывная и органическая связь. Проходит не так уж много времени, — и теория становится практикой в связи с разработкой машинного перевода.

В очерке «Формулы фонемы» мы рассказывали об Иване Александровиче Бодуэне де Куртене, о том, как его идеи о фонеме, благодаря преемственности русской фонологической школы, завоевали весь мир и фонология стала ведущей лингвистической дисциплиной.

Вот другой пример преемственности в отечественной лингвистической науке. Академик Филипп Федорович Фортунатов в конце прошлого столетия создал так называемую московскую лингвистическую школу. Основной упор Фортунатов делал на грамматическое описание языка, стремясь сделать его предельно строгим и четким. Идеи эти были развиты в нашем веке профессором Александром Матвеевичем Пешковским в его капитальном труде «Русский синтаксис в научном освещении» (первое издание вышло в 1914 г., последнее прижизненное издание — в 1928 г.). Тридцать лет спустя именно в нашей стране начинается описание грамматики языка с помощью математической логики и теории множеств. Математические модели такого описания предлагают математики А. Н. Колмогоров, Р. Л. Добрушин, идеи эти развиваются математиком и программистом О. С. Кулагиной, а лингвист И. И. Ревзин выпускает первую в мире монографию, посвященную теории моделей языка, вызвавшую широкий отклик у советских и зарубежных математиков, логиков и лингвистов.

Иногда мы имеем дело не с прямой преемственностью идей и поколений, а со взглядами людей, которые на несколько десятков лет, а порой и на целые столетия опережали свою эпоху. Вспомним Раймонда Луллия и

его вертушку, о которой мы рассказывали в очерке «МП, ЯП, ИЯ». Луллий жил в XIII столетии. Идеи его были развиты жившим в XVII веке Лейбницем, который был не только гениальным философом, математиком и логиком, пытавшимся создать «алгебру языка» и на ее основании «алгебру мысли», но и великим изобретателем — он сконструировал первый в мире арифмометр. Но только в нашем веке, благодаря электронике, идеи Луллия, Лейбница и других ученых, изобретателей получили реальную почву для своего технического воплощения, для создания ЭВМ.

Здесь мы переходим к еще одной давней традиции лингвистики — к ее неразрывной связи с практикой, с жизнью. Все мы лингвисты поневоле, когда обучаемся родному языку и грамоте, когда изучаем иностранные языки и овладеваем правильной литературной речью. В наши дни на помощь приходят радио и телевидение, магнитофоны и пластинки, сосуществующие с традиционными словарями, грамматиками и учебниками. В очерке «Жар холодных чисел» мы рассказали о том, как современная лингвистика помогает процессу обучения, делает его более быстрым и надежным благодаря частотным словарям, машинному сравнению учебников иностранных языков. Нет сомнения в том, что в будущем эта помощь станет еще более эффективной и многосторонней.

Но какие бы частотные словари ни были созданы, сколько бы языков ни изучил человек с помощью новых методов обучения, ему все равно не справиться с тем потоком информации, который обрушивается на его голову. В наши дни за одну минуту в мире появляется две тысячи страниц научно-технической информации, и для того, чтобы следить только за литературой по своей узкой специальности, современному ученому нужно в день прочитывать до полутора тысяч страниц на разных языках мира, причем к традиционным европейским языкам в последнее время прибавились публикации на хинди, арабском, японском языках, пользующихся оригинальными системами письма. Ясно, что без помощи ЭВМ обуздать этот поток информации невозможно. И столь же ясно, что ЭВМ должна получить программу действий от современной лингвистической науки.

Речь идет не только о машинном переводе и создании информационно-поисковых систем. Не случайно в

статье, опубликованной в первом номере журнала «Коммунист» за 1977 год, подчеркивалось, что «использование ЭВМ на предприятиях невозможно без моделирования многих важных, хотя пока и сравнительно простых мыслительных операций, решения проблем перевода с одного языка на другой — языка, понятного человеку, на язык, «понятный» машинам». Чем дальше, тем сложнее будут становиться эти операции, по мере того как мы будем все глубже и глубже постигать тайны нашего языка.

Когда-то журналисты с восторгом писали о машине «Урал», выполняющей сто операций в секунду и хранящей в памяти около тысячи чисел. В наши дни созданы ЭВМ, работающие со скоростью сотен миллионов операций в секунду и хранящие в памяти до миллиарда чисел. Ученые всерьез обсуждают вопрос о создании машин, память которых будет ассоциативной, многомерной, то есть максимально приближенной к человеческой.

Но какими бы быстродействующими ни были ЭВМ, какой бы чудовищной памятью они ни обладали и сколько бы ассоциативных связей эта память ни вмещала, создание «электронного мозга» невозможно до тех пор, пока наука о языке не решит своих кардинальных проблем. Ибо именно на этом языке человечество закодировало все основные знания, накопленные за свою многотысячелетнюю историю.

«Может ли машина мыслить?» — в дискуссии на эту тему выступили философы и логики, математики и психологи, социологи и программисты. Свой авторитетный взгляд на возможность создания «мыслящих машин» высказали ученые, чьи работы легли в фундамент кибернетики: Норберт Винер, А. Н. Колмогоров, А. Тьюринг. Проблема эта широко освещалась на страницах журнала «Вопросы философии» и в различных научных и научно-популярных изданиях. Дискуссия, начавшаяся в 50-х годах, на заре кибернетики, и столь актуальная в 60-е годы, осталась незавершенной. А ныне специалисты ведут разговор о создании систем искусственного интеллекта. Они не спорят о принципиальных возможностях кибернетических машин и отличиях их от мозга человека, а обсуждают конкретные работы, реальные программы и автоматы, по этим программам действующие. Проблемам искусственного интеллекта посвяща-

ются научные симпозиумы в нашей стране и международные конференции. Лаборатории искусственного интеллекта существуют во многих крупнейших исследовательских центрах мира. Отказавшись «объять необъятное», промоделировать работу человеческого мозга во всей ее полноте и широте, ученые наших дней создают и реализуют программы, благодаря которым ЭВМ выполняют функции, до сих пор считавшиеся доступными только разуму людей. Например, формируют новые понятия, строят и проверяют гипотезы, принимают решения в сложных, меняющихся ситуациях и т. п.

Мы могли бы назвать еще целый ряд отечественных и зарубежных систем искусственного интеллекта, как уже работающих, так и проектируемых. Современные ЭВМ с их фантастическим быстродействием представляют колоссальные возможности для таких систем. Главной же трудностью в их создании и совершенствовании является все тот же человеческий язык, проблема его формализации, перевода на язык машины. «Устройства, способные выполнять «интеллектуальные действия», призваны заменить человека в тех случаях, когда его прямое участие невозможно или нецелесообразно,— пишет В. В. Мартынов в статье «Семиологические проблемы искусственного интеллекта», открывающей первый выпуск «Известий Академии наук СССР» за 1978 год.— Поскольку единственное назначение подобных устройств заключается в том, чтобы они служили человеку, необходимо обеспечить их регулярное общение с человеком. Таким образом, центральной проблемой искусственного интеллекта оказывается проблема языка или языков, с помощью которых могли бы вести диалог с человеком интеллектуальные устройства».

Решить эту центральную проблему, проблему языка, разумеется, без специалистов по лингвистике нельзя. Ни программист, ни логик, ни психолог не заменят здесь языковеда, который знает уже многие тонкости и сложности языка и который может указать, что из них необходимо учитывать при решении той или иной задачи, а чем можно и пренебречь.

Лингвистическим проблемам искусственного интеллекта посвящены многие работы языковедов и кибернетиков, эта тема обсуждалась на специальной конференции в Ленинграде, и нет сомнения в том, что дальнейшие успехи «интеллектуальных машин» в значитель-

ной мере зависят от успехов современного языкознания. Однако и сама лингвистика приобретает от такого содружества необычайно много. Союз с кибернетикой позволяет лингвистике выйти за рамки «описания языка через язык», образующего замкнутый круг (язык описывается нами также на языке, описание это требует своего языка и т. д. до бесконечности).

Лингвистическим автоматом называют устройство, способное воспринимать, перерабатывать и выдавать информацию, закодированную на языке человека. Это машина-переводчик, машина — составитель словарей, машина, реферирующая тексты... Но только ли для автоматов нужно разрабатывать лингвистические программы и модели? Ведь автомат, включая самые совершенные ЭВМ, — это, по сути дела, лишь ящик для переработки информации, «голый мозг», лишенный органов чувств и средств передвижения. И, что самое существенное, автомат не обладает тем, что есть у каждой человеческой личности, — моделью мира, своеобразным, но обязательным «фильтром», через который проходит получаемая информация. Человеческий язык неразрывно связан с мышлением и мировоззрением людей, он создан в обществе и для общества, это универсальное и уникальное орудие познания окружающего мира, это «орудие мысли» (вспомните наш очерк «Модель мира»).

Помимо «электронного мозга» ЭВМ, созданы машины, способные манипулировать предметами, ориентироваться в искусственной, экспериментальной обстановке, передвигаться в пространстве. Роботы эти наделяются электронным «глазом», а действиями их руководит «мозг» вычислительной машины. В память робота закладывается определенная программа, формирующая внутреннюю «модель мира», пускай пока что и очень примитивную. Вот для таких устройств и разрабатываются специальные языковые программы (направление это, возникшее на стыке языкознания, программирования и роботостроения, получило наименование «лингвистика для роботов», или «роботолингвистика»).

В этом направлении сделаны пока что лишь первые шаги. Так, в нашей стране вышла монография Т. Винограда под названием «Программа, понимающая естественный язык», где описывается робот, выполняющий приказания человека, даваемые не на языке алго-

ритмов, а на обычном языке (правда, внутренний мир робота был очень ограничен, а действия он производил с помощью манипулятора над тремя типами геометрических фигур, окрашенных в три разных цвета). Но даже эти первые эксперименты позволяют нам сделать еще один шаг к самому сокровенному в человеческом языке, к постижению его связи с деятельностью и мышлением.

В различных языках мира существуют разнообразнейшие системы времен. Вспомним хотя бы английские глаголы, чья временная система не совпадает с системой времен русского языка. Кибернетики, работающие в области систем искусственного интеллекта, создают универсальную систему времен для роботов, в которой любое событие имеет определенные координаты на «оси времен». Совмещение «человеческих» языковых времен и логического «времени робота» приносит пользу не только программистам, но и лингвистам. «Лингвистика для робота» помогает ученым лучше и четче осознать и наш собственный язык, его возможности отражать события во времени. То же самое можно сказать и о категории пространства (математики создали не только временную логику, частный случай логики модальной, но и специальные пространственные логики — метрическую, релятивную, топологическую).

Насколько же совпадают и в чем принципиально расходятся пространственно-временные «представления» робота, созданные на основе новейших разделов математики, от представлений людей, у которых в зависимости от родного языка существуют свои «национальные» координаты времени и пространства? Где границы, что разделяют «лингвистику для роботов», действующих по заданным моделям мира, от законов языка, которым пользуемся мы, живые люди, мыслящие существа?

Вопросы эти уже поставлены. Будем ждать ответа.

## СОДЕРЖАНИЕ

Р. Г. Пиотровский. Двенадцать лет спустя. (Вместо предисловия) . . . . .	5
Жар холодных числ... . . . . .	9
МП, ЯП, ИЯ . . . . .	43
В поисках значения . . . . .	72
Наш удивительный код. . . . .	98
Модель мира . . . . .	124
Формулы фонемы . . . . .	146
Черный ящик . . . . .	172
1878—1978 и далее... (Вместо послесловия) . . .	196

**Александр Михайлович КОНДРАТОВ**

**ЗВУКИ И ЗНАКИ**

Зав. редакцией научно-художественной литературы

**М. Новиков**

Редактор **Н. Яснопольский**

Мл. редактор **В. Саморига**

Оформление **М. Дорохова**

Худож. редактор **М. Гусева**

Техн. редактор **Т. Луговская**

Корректор **Н. Мелешкина**

**ИБ № 1350**

А 04420. Индекс заказа 87729. Сдано в набор 29.03.78 г.  
Подписано к печати 24.10.78 г. Формат бумаги 84×108<sup>1</sup>/<sub>32</sub>.  
Бумага типографская № 1. Бум. л. 3,25. Печ. л. 6,5. Усл.  
печ. л. 10,92. Уч.-изд. л. 11,14. Тираж 95 000 экз. Издательство  
«Знание». 101835, Москва, Центр, проезд Серова, д. 4.  
Заказ 958. Цена 45 коп.

Киевская книжная фабрика республиканского производствен-  
ного объединения «Полиграфкнига» Госкомиздата УССР,  
Киев, ул. Воровского, 24.