4
П-78

АКАДЕМИЯ НАУК СССР

ПРОБЛЕМЫ СТРУКТУРНОЙ ЛИНГВИСТИКИ

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР

1 ОЯ - сборн. сущ. содерж.
2. Перед. мысль. частями -
передн. отнс. и сзади
заключение

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

57859

АКАДЕМИЯ НАУК СССР
ИНСТИТУТ РУССКОГО ЯЗЫКА

[illegible]
27853

ПРОБЛЕМЫ СТРУКТУРНОЙ ЛИНГВИСТИКИ

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР
Москва 1962

4
П-78
БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
ИН-ТА ЯЗЫКОЗНАНИЯ АН СССР
57859

Ответственный редактор
С. К. ШАУМЯН

ПРЕДИСЛОВИЕ

Структурную лингвистику можно определить в самом общем виде как науку, имеющую своим предметом изучение естественных языков с точки зрения их преобразования в абстрактные коды, служащие формальными моделями естественных языков.

Лингвистическое моделирование, представляющее специфическую задачу структурной лингвистики, осуществляется применительно ко всем аспектам изучения языка — фонологии, грамматике, лексикологии и, наконец, при изучении письменного облика языка — в графемике и теории орфографии.

Лингвистическое моделирование имеет не только фундаментальное теоретическое значение, но и первостепенную важность для решения проблем прикладной кибернетики в области автоматизации поиска научно-технической информации, речевого управления производственными объектами, конструирования автоматических читающих и стенографирующих устройств, повышения пропускной способности каналов проводной и непроводной связи и в ряде других областей.

Современное состояние лингвистического моделирования характеризуется следующими существенными чертами.

1. Поскольку всякое, в том числе и лингвистическое, моделирование требует применения математических методов, в структурную лингвистику все более и более интенсивно внедряются математические методы. Понятие языка, которым оперирует структурная лингвистика, — это язык в соссюровском понимании — язык, обладающий нестатистической структурой. Поэтому фундаментом построении лингвистических моделей должны быть именно нестатистические математические методы — для структурной лингвистики прежде всего важны методы математической логики, теории множеств, теории графов. Нестатистические лингвистические модели могут быть названы базисными лингвистическими моделями. Так как язык функционирует через речь, для которой характерна статистическая структура, то от базисных лингвистических моделей естествен переход к разного рода статистическим и вероятностным моделям. Для построения такого рода моделей необходимы методы математической статистики, теории вероятностей, теории информации. Между базисными лингвистическими моделями и моделями речи существует иерархия: все эффективные модели речи должны опираться на базисные лингвистические модели, служащие непосредственным объектом структурной лингвистики.

2. Как во всякой другой теоретической науке, в структурной лингвистике существенное значение имеет исследование отношения моделей к отражаемой ими реальности, в данном случае — отношение лингвистической модели к лингвистической реальности. Но для подобного рода исследования необходим выход за пределы собственной области данной теоретической науки и обращение к общим проблемам логики науки. Поэтому для современного развития структурной лингвистики чрезвы-

чайно важна ее связь с логикой науки. Поскольку за последнее время целый ряд фундаментальных проблем, связанных с процессом познания, изучается в теоретической кибернетике (здесь нужно упомянуть прежде всего кибернетическую теорию черного ящика), то, помимо уже упомянутой выше связи структурной лингвистики с прикладными областями кибернетики, вполне естественна связь структурной лингвистики с кибернетикой и в общеметодологическом плане, в плане исследования процесса лингвистического познания под кибернетическим углом зрения.

3. Для всякого моделирования, в том числе и для лингвистического моделирования, существенно многообразие подходов. Всякая модель, как бы она ни была совершенна, все же не может не быть односторонней в силу своей абстрактной природы. Поэтому для эффективного отражения реальности необходимы разнообразные модели, разнообразные подходы к конструированию моделей. Для современного состояния структурной лингвистики характерно именно многообразие точек зрения и подходов к конструированию моделей, и это обстоятельство является залогом ее дальнейших успехов.

Указанные черты, характеризующие современное состояние и тенденции развития структурной лингвистики, в той или иной степени нашли отражение и в статьях настоящего сборника.

Прежде всего укажем, что в сборнике отражены различные точки зрения и подходы к конструированию лингвистических моделей. Вместе с тем необходимо отметить и внутреннее единство сборника, которое определяется спецификой структурной лингвистики как науки о лингвистическом моделировании: идет ли речь об общих проблемах структурной лингвистики, или о проблемах фонологии, структурной грамматики, структурной лексикологии или графемики — всюду красной нитью проходит идея лингвистического моделирования на различных уровнях или под различными углами зрения. Все статьи сборника, за исключением статьи Р. Г. Пиотровского о теоретико-информационном изучении речи, посвящены базисному лингвистическому моделированию, являющемуся непосредственным объектом структурной лингвистики. Мы сочли целесообразным включить в настоящий сборник интересную статью Р. Г. Пиотровского, чтобы показать естественную связь между базисным лингвистическим моделированием и моделированием речи.

В сборнике читатель найдет приложение к лингвистическому моделированию идей теории множеств, теории графов, математической логики, обсуждение теоретических проблем с точки зрения общих идей логики науки и кибернетики.

С. К. Шаумян

I. ОБЩИЕ ПРОБЛЕМЫ

С. К. Шаумян

ПРЕОБРАЗОВАНИЕ ИНФОРМАЦИИ В ПРОЦЕССЕ ПОЗНАНИЯ И ДВУХСТУПЕНЧАТАЯ ТЕОРИЯ СТРУКТУРНОЙ ЛИНГВИСТИКИ

С точки зрения кибернетики всякий объект познания может рассматриваться в качестве так называемого черного ящика.

Понятие черного ящика заимствовано из электротехники. В электротехнике часто бывает необходимо определить содержимое опечатанного ящика, имеющего входные и выходные зажимы. Подвергая входные зажимы разным воздействиям и наблюдая результаты этих воздействий на выходных зажимах, инженер делает заключение о содержимом ящика. Подобные ящики, содержимое которых недоступно прямому наблюдению, называются черными ящиками.

Если рассматривать объект познания в качестве эпистемологического черного ящика, то его входом будут операции, производимые экспериментатором над объектом, а его выходом наблюдения над результатами этих операций.

Объект познания и экспериментатор представляют собой систему с обратной связью. Экспериментатор соединяет себя с черным ящиком, подвергая черный ящик разного рода воздействиям и наблюдая за результатами этих воздействий. На основе полученных результатов черный ящик подвергается новым воздействиям, которые дают новые результаты. Таким образом, экспериментатор действует на черный ящик, а черный ящик действует на экспериментатора, и в результате этого взаимодействия происходит углубление познания. Основу науки составляет так называемый протокол, под которым имеется в виду фиксирование состояний входа (операции над исследуемым объектом) и выхода (результаты операций над исследуемым объектом) эпистемологического черного ящика. Информация, доставляемая протоколом, служит исходной базой для раскрытия сущности недоступных прямому наблюдению элементов и глубинных связей между этими элементами.

Задача теоретического исследования заключается в том, чтобы преобразовать информацию, доставляемую протоколом, в информацию о недоступной прямому наблюдению сущности познаваемого объекта. Информацию первого рода мы называем элементарной информацией, а информацию второго рода — информацией о конструктах. Таким образом, жизненным нервом процесса познания должно считаться перекодирование элементарной информации в информацию о конструктах.

Информация о конструктах представляется в виде формальной модели, но оба эти понятия вовсе не тождественны друг другу: хотя информация о конструктах всегда должна представляться в виде формальной модели, но обратное не верно, не всякая формальная модель содержит информацию о конструктах, существуют также формальные модели, которые служат

аналогами прямо наблюдаемых факторов и явлений. К последним относятся модели, опирающиеся на статистические методы.

С кибернетической точкой зрения на процесс познания вполне гармонирует точка зрения современной логики науки, различающая в структуре науки две главных ступени абстракции — ступень наблюдения и ступень конструктов [1].

Кибернетическая точка зрения на процесс познания имеет фундаментальное значение для структурной лингвистики, как и для остальных наук. С точки зрения кибернетики соссюровское противопоставление языка и речи можно трактовать так: речь есть совокупность входных и выходных состояний лингвистического черного ящика, а язык есть недоступное прямому наблюдению содержимое лингвистического черного ящика.

Процесс лингвистического исследования можно трактовать так: исследователь подвергает лингвистический черный ящик разным операциям на входе и наблюдает результаты этих операций на выходе; полученную таким образом элементарную лингвистическую информацию исследователь преобразует в информацию о лингвистических конструктах.

Если обозначить язык как предмет исследования термином «язык-объект», то элементарную информацию о языке-объекте мы будем называть термином «метаязык первой степени», а информацию о лингвистических конструктах мы будем называть термином «метаязык второй степени». Таким образом, лингвист должен оперировать двумя метаязыками. Задача состоит в том, чтобы путем применения соответствующих операций прежде всего фиксировать метаязык первой степени и затем на основе теоретического анализа метаязыка первой степени преобразовать его в метаязык второй степени.

Можно провести некоторую аналогию между работой лингвиста-теоретика и работой переводчика с одного языка на другой. В обоих случаях мы имеем дело с перекодированием информации, то есть с переводом информации из одного кода в другой код. При этом, однако, есть и серьезное различие между обоими видами работы. В отличие от обыкновенного перевода с одного языка на другой, перевод с метаязыка первой степени на метаязык второй степени является своеобразным переводом-рефератом, при котором сохраняются только существенные черты оригинала, а все несущественные черты опускаются.

Говоря об отличии работы лингвиста-теоретика от работы переводчика с одного языка на другой, следует обратить внимание еще и на другую сторону дела. Как известно, процесс перевода можно представить в виде строгих алгоритмических правил. Эти алгоритмические правила механизируют процесс перевода таким образом, что он может осуществляться машиной. В противоположность переводу с одного языка на другой, преобразование метаязыка первой степени, описывающего данные ступени наблюдения, в метаязык второй степени, относящийся к ступени конструктов, не может быть сведено к системе механических алгоритмических правил. Характеризуя конструкты в физике, А. Эйнштейн и Л. Инфельд пишут: «Физические понятия суть свободные творения человеческого разума и не однозначно определены внешним миром, как это иногда может показаться» [2]. Точно так же обстоит дело с конструктами в других науках, в том числе в науке о языке. Лингвистические

---

[1] Подробное изложение теории черного ящика читатель может найти в книге: У. Росс Эшби. Введение в кибернетику. М., 1959, стр. 127—169; о двух ступенях абстракции в структуре науки см.: R. Carnap. The methodological character of theoretical concepts. In: «Minnesota studies in the philosophy of science», vol. I. Minneapolis, 1956.

[2] А. Эйнштейн, Л. Инфельд. Эволюция физики, изд. 2-е. М., 1956, стр. 62.

конструкты — это, говоря словами А. Эйнштейна и Л. Инфельда, свободные творения человеческого разума. Поэтому они не выводятся из данных наблюдения путем статистических или каких-либо других механических правил, а просто постулируются в целях объяснения этих данных. Таким образом, мы можем утверждать, что в науке о языке преобразование метаязыка первой степени в метаязык второй степени сводится к постулированию метаязыка второй степени для объяснения метаязыка первой степени.

Двухступенчатая теория структурной лингвистики, к изложению которой мы приступаем, находится в полном согласии с только что изложенным пониманием структуры науки в кибернетике и в современной логике науки. Это, конечно, никоим образом не означает, что двухступенчатая теория структурной лингвистики возникла, так сказать, «сверху», то есть путем априорного наложения на структурную лингвистику логической схемы структуры науки. Мы должны со всей силой подчеркнуть, что разграничение двух ступеней абстракции в структурной лингвистике может и должно быть объяснено с точки зрения современной логики, но само по себе оно идет не от логики, а от лингвистических фактов. Разграничение двух ступеней абстракции в структурной лингвистике диктуется не общелогическими соображениями, а необходимостью преодолеть фундаментальные теоретические трудности, фундаментальные антиномии, возникающие при анализе лингвистических фактов.

В фонологии мы имеем в виду антиномию, которую мы называем антиномией транспозиции. К анализу этой антиномии мы и переходим.

К числу основных тезисов, на которые опирается современная фонология, принадлежат:

1) фонемы есть элементы, служащие для различения лингвистических единиц;

2) фонемы есть акустические элементы.

Признавая эти тезисы истинными, мы должны считать истинными и все следствия, вытекающие из этих тезисов.

Начнем с тезиса № 1.

Если фонемы есть элементы, служащие для различения лингвистических единиц, то отсюда следует, что мы можем траспонировать фонемы из акустической субстанции в другие виды субстанций (в графическую, цветовую, тактильную и т. д.). Например, каждую фонему можно транспонировать в кружок определенного цвета; в этом случае план выражения будет представлять собой поток следующих друг за другом кружков разного цвета, служащих для различения лингвистических единиц.

Все исторически сложившиеся естественные языки — звуковые. Поэтому с точки зрения естественных языков основная субстанция — акустическая, все остальные виды субстанции подчинены ей. Однако данная констатация эмпирической неравноправности разного рода субстанций не затрагивает коммуникативной сущности языка. Если стать на точку зрения коммуникативной сущности естественных языков, то надо признать полную равноправность всех видов субстанции. С этой точки зрения, скажем, графическая субстанция является столь же основной, как и акустическая. В связи с этим очень интересно замечание Б. Рассела о том, что в нашем распоряжении нет данных, позволяющих решить, какая форма языкового выражения является более древней — звуковая или письменная [3]. В самом деле, если с точки зрения коммуникативной сущности языка акустическая и графическая субстанции равноправны, то вполне допустима гипотеза, что в древнейшие эпохи основной субстанцией выражения естественных языков была графическая. Заметим мимоходом, что не следует смешивать элементы графической субстанции

[3] См.: B. Russell. An outline of philosophy. London, 1927, стр. 47.

с буквами алфавита. В зависимости от степени совершенства той или иной системы орфографии буквы алфавита с большим или меньшим успехом используются в качестве символов фонем (конечно, наряду с функцией обозначения фонем буквы алфавита могут обладать и, как правило, обладают морфологическими и другими лингвистическими функциями). Что же касается элементов графической субстанции, то, в отличие от букв, это не знаки, а графические диакритические элементы, служащие аналогами акустических диакритических элементов.

Если из тезиса № 1 следует, что мы можем транспонировать фонемы из акустической субстанции в другие виды субстанции, то из тезиса № 2 следует прямо противоположное. В самом деле, если верно, что фонемы есть акустические элементы, то мы не вправе транспонировать фонемы из акустической в другие виды субстанции.

Выходит, что фонема может и не может транспонироваться из акустической субстанции в другие виды субстанции.

Такова фундаментальная антиномия, на которую наталкивается анализ лингвистических фактов плана выражения.

Мы полагаем, что преодолеть данную антиномию можно только в том случае, если мы расщепим понятие фонемы на понятие фонемы в собственном смысле и понятие фонемоида как физической субстанции, служащей субстратом фонемы. Фонемы — это идеальные диакритические элементы, которые могут транспонироваться из одной физической субстанции в другую. Что же касается фонемоидов, то это — воплотители фонем, то есть элементы, без которых фонемы сами по себе существовать не могут.

В современной фонологии фонемы рассматриваются как пучки дифференциальных элементов.

В связи с расщеплением понятия фонемы на понятие фонемы как идеального диакритического элемента и понятие фонемоида возникает необходимость подвергнуть пересмотру и понятие дифференциального элемента.

Относительно дифференциальных элементов общеприняты два тезиса:

1) дифференциальные элементы есть не разложимые далее элементы, служащие для различения лингвистических единиц;

2) дифференциальные элементы есть акустические элементы.

Рассмотрим следствия, вытекающие из обоих тезисов.

Если верно, что дифференциальные элементы служат для различения лингвистических единиц, то отсюда следует, что существует принципиальная возможность транспонирования акустической субстанции в другие виды физической субстанции — в графическую, цветовую, тактильную. Любую систему дифференциальных элементов мы можем представить не только в виде акустических свойств, но и в виде графических, цветовых и тактильных символов.

Для доказательства принципиальной возможности транспонирования акустической субстанции в другие виды физической субстанции проведем следующий мысленный эксперимент. Транспонируем дифференциальные элементы в квадраты одинакового размера, но разного цвета, скажем, вокальность в темно-синий квадрат, невокальность — в светло-синий, консонантность — в темно-коричневый квадрат, неконсонантность — в светло-коричневый, компактность — в темно-зеленый, диффузность — в светло-зеленый, низкую тональность — в темно-желтый квадрат, высокую — в светло-желтый квадрат и т. д. Тогда каждую фонему можно представить в виде комбинации квадратов разного цвета. Например, если взять систему гласных фонем русского языка, то в рамках приведенной выше панхронической системы дифференциальных элементов фонему *а* можно представить в виде комбинации темно-синего, светло-коричневого и темно-зеленого квадратов, фонему *і* — в виде комбинации квадратов темно-синего, светло-коричневого, светло-зеленого и светло-желтого, фонему *е* —

в виде комбинации квадратов темно-синего, светло-коричневого, темно-зеленого, светло-зеленого и светло-желтого, фонему *о* — в виде комбинации квадратов темно-синего, светло-коричневого, темно-зеленого, светло-зеленого и темно-желтого, фонему *и* в виде комбинации квадратов темно-синего, светло-коричневого, светло-зеленого и темно-желтого.

Итак, из тезиса № 1 следует, что дифференциальные элементы могут транспонироваться из акустической субстанции в другие виды физической субстанции.

Обратимся теперь к тезису № 2. Если верно, что дифференциальные элементы есть акустические элементы, то отсюда следует, что они не могут транспонироваться в другие виды физической субстанции, иначе они перестанут быть самими собой, то есть акустическими элементами.

Налицо явная антиномия: и тезис № 1 и тезис № 2 признаются верными современной фонологией, но из тезиса № 1 следует, что дифференциальные элементы могут транспонироваться из акустической в другие виды физической субстанции, а из тезиса № 2 следует прямо обратное — что мы не можем транспонировать дифференциальные элементы из акустической в другие виды физической субстанции.

Единственный путь для преодоления этой антиномии — расщепить понятие дифференциального элемента на понятие дифферентора и понятие дифферентоида. Дифференторы — это идеальные дифференциальные элементы, которые могут транспонироваться из одной физической субстанции в другую. Дифферентоиды — это воплотители дифференторов, то есть элементы, которые обеспечивают функционирование дифференторов как их физические субстраты.

Принципиальное отличие фонем и дифференторов от фонемоидов и дифферентоидов заключается в том, что фонемы и дифференторы не наблюдаются нами, а постулируются в качестве теоретических понятий для объяснения прямо наблюдаемых факторов.

Представив только что изложенную систему фонологических понятий, мы можем теперь объяснить ее с точки зрения логического принципа двухступенчатости. В свете современной логики науки фонемоиды и дифферентоиды относятся к эмпирической базе науки, а фонемы и дифференторы к уровню так называемых конструктов (конструктами современная логика науки называет понятия о ненаблюдаемых объектах). Однако разграничение двух ступеней абстракции в фонологии идет не от логики, а от анализа фонологических фактов самих по себе. Логика только позволяет осмыслить этот анализ с точки зрения логической теории познания[1].

Перейдем теперь к анализу основной антиномии в грамматике. Мы имеем в виду антиномию формального анализа.

Краеугольным камнем современной структурной лингвистики служит следующий тезис: лингвистический анализ должен быть строго формальным. Вместе с тем структурная лингвистика, как и всякая наука, опирается на общеметодологический тезис, что во всякой научной системе недопустимы взаимоисключающие высказывания. Однако последовательное проведение формального анализа приводит именно к взаимоисключающим высказываниям. Возьмем, например, в русском языке фразу *Посещение ее братом учителя*. Со строго формальной точки зрения эта фраза допускает две взаимоисключающие интерпретации: либо фраза *посещение ее* определяется фразой *братом учителя*, либо фраза *посещение учителя* определяется фразой *ее братом*. Обе эти интерпретации с формальной точки зрения в равной степени правомерны, хотя они исключают друг друга. Другое дело, если мы прибегнем к анализу значения: в зави-

---

[1] Подробное систематическое изложение предлагаемой фонологической теории читатель найдет в книге: С. К. Шаумян. Проблемы теоретической фонологии. М., Изд. АН СССР, 1962.

симости от значения фразы мы можем сделать формальный анализ мотивированным. Однако критерий значения явится в подобных случаях приемом ad hoc, нарушающим принцип гомогенности. Нам представляется, что единственный путь для преодоления данной антиномии заключается в том, чтобы расщепить основные понятия грамматики — морфему и синтагму — на понятия морфемы и синтагмы как чисто реляционных элементов и понятия морфемоида и синтагмоида как реляционно-семантических элементов. При таком расщеплении речь будет идти не о случайном применении критерия значения в качестве приема ad hoc, а о систематическом анализе реляционно-семантических единиц на уровне наблюдения и построении чисто формальной системы на уровне конструктов.

Для того, чтобы показать необходимость расщепления морфемы на морфему$_1$ и морфемоид и синтагмы на синтагму$_1$ и синтагмоид (цифровой индекс у терминов «морфема$_1$» и «синтагма$_1$» показывает, что речь идет о конструктах в отличие от общепринятого понимания терминов «морфема» и «синтагма»), прибегнем к трансформационному анализу.

Если применить трансформационный анализ к фразе *посещение ее братом учителя*, то можно получить две трансформации:

1) *Посещение ее братом учителя* < *Ее брат посетил учителя*
2) *Посещение ее братом учителя* < *Брат учителя посетил ее*

Таким образом, мы получаем возможность выбирать между двумя формальными интерпретациями фразы «посещение ее братом учителя» в зависимости от того, с какой трансформацией мы эту фразу соотносим. В связи с этим возникает следующий вопрос: в какой мере трансформационный анализ обеспечивает последовательное проведение формального принципа в структурной лингвистике? Не приводит ли трансформационный анализ к смешению формальной и семантической точек зрения?

Чтобы разобраться в этом вопросе, сопоставим следующие две фразы: *Учитель молод*, *Учитель рад*.

Поскольку именная синтагма может рассматриваться в качестве редуцированной предикативной синтагмы, то мы можем постулировать трансформацию [5]: *Учитель молод* > *Молодой учитель*.

А как быть с синтагмой *Учитель рад*?

Если рассматривать данную предикативную синтагму в ее семантической конкретности, то ее невозможно редуцировать в соответствующую именную синтагму. Но в таком случае мы не в праве считать трансформационный анализ формальным методом, как на этом настаивает Н. Хомский и другие исследователи.

Именно для преодоления данной и аналогичных трудностей, связанных с применением трансформационного анализа, мы и считаем необходимым расщепить понятия морфемы и синтагмы на понятия морфемы$_1$ и морфемоида и синтагмы$_1$ и синтагмоида.

Если осуществить данное расщепление понятий, то становится ясным, что фраза *Учитель рад* так же подлежит трансформации, как и фраза *Учитель молод*, поскольку трансформируются абстрактные синтагмы-конструкты, а не синтагмоиды, служащие реляционно-семантическими субстратами синтагм-конструктов.

Чтобы принципиальное различение между синтагмами, морфемами и синтагмоидами, морфемоидами было более осязательно, целесообразно проанализировать некоторые фразы с помощью символической записи.

---

[5] Можно и, наоборот, рассматривать предикативную синтагму как развернутую именную синтагму. Поскольку для нашей проблемы несущественно, какой из этих подходов считать более правильным, то, чтобы не отвлекаться от существа рассматриваемой проблемы, мы будем опираться на общепринятую точку зрения, согласно которой трансформация направлена от предикативной синтагмы к именной.

Введем следующие символические обозначения:

$\rightarrow$ — отношение детерминации;

$\leftrightarrow$ — отношение интердепенденции;

$x \leftrightarrow y$ — предикативная фраза;

$x \rightarrow y$ — именная фраза;

Rd — оператор редукции предикативной фразы в именную фразу;

“ ” (кавычки) — знак для обозначения морфем в отличие от морфемоидов.

Например, если a, b, c — морфемоиды, то “a“, “b“, “c“ — соответствующие морфемы.

Возьмем теперь фразу *Прибытие поезда* и проанализируем ее на двух ступенях абстракции. Если обозначить слово *поезда* знаком a, а слово *прибытие* знаком b (для упрощения записи мы приравниваем слова к простым элементам), то на ступени наблюдения данную фразу можно представить в виде формулы: $a \rightarrow b$. Но на ступени конструктов мы будем иметь принципиально отличную формулу: Rd(“a“$\leftrightarrow$“b“).

Первая формула обозначает синтагмоид, а вторая формула — соответствующую синтагму. Вторая формула есть итог трансформации синтагмы, субстратом которой служит фраза *поезд прибыл* в синтагму, субстратом которой служит фраза *прибытие поезда*.

Сопоставим теперь следующие две фразы: *Пение артиста*, *Чтение книги*. На уровне прямо наблюдаемых отношений между обеими фразами нет никакой разницы, поскольку в обоих случаях синтагмоиды выражаются тождественной формулой $a \rightarrow b$. Если же подняться на уровень конструктов, то мы обнаружим, что одинаковые синтагмоиды служат субстратами разных синтагм. В первом случае синтагма будет иметь формулу Rd (“a“$\leftrightarrow$“b“), а во втором случае формулу Rd [“c“$\leftrightarrow$(“b“ $\rightarrow$ “a“)]

Первая формула есть итог трансформации

Артист поет $\rightarrow$ *Пение артиста*

Вторая формула есть итог трансформации

Кто-то читает книгу $\rightarrow$ *чтение книги* (Во второй формуле “c“ соотносится с *кто-то*, “b“ — с *читает*, “a“ — с *книгу*).

Приведенные примеры символической записи наглядно показывают принципиальное качественное различие между синтагмами, морфемами и синтагмоидами, морфемоидами.

В связи с различением двух ступеней абстракции мы считаем нецелесообразным говорить о трансформациях применительно к синтагмоидам. С нашей точки зрения трансформироваться могут только синтагмы. Что же касается синтагмоидов, то они сами по себе не трансформируются, но могут ограничивать сферу применения тех или иных трансформаций. Возьмем, например, английскую фразу *He is asleep*, приведенную в работе Р. Лиза [6]. По мнению Р. Лиза, данная предикативная синтагма не подлежит трансформации в именную синтагму, потому что в английском языке невозможна фраза *The asleep man*. Мы полагаем, что поскольку фраза *He is asleep* в ее конкретной данности представляет собой синтагмоид, то применительно к этой фразе нельзя ставить вопрос о возможности или невозможности трансформации. Даже если бы фраза *The asleep man* была возможна в английском языке, все равно и в этом случае нельзя было бы говорить о трансформации фразы *He is asleep*. Суть дела заключается в том, что фраза *He is asleep*, как и другие аналогичные фразы английского языка, служит субстратом предикативной синтагмы “a“$\leftrightarrow$“b“, которая всегда может трансформироваться.

Законы лингвистических трансформаций, как и любые другие законы действительности, рассматриваемой на ступени конструктов, могут иметь более или менее широкую сферу своего действия, но не могут допускать

---

[6] См.: R. B. Lees. The grammar of English nominalizations. «International Journal of American Linguistics», vol. XXVI, № 3, part II, 1960, стр. 16.

исключений. С исключениями мы можем иметь дело только на уровне эмпирической генерализации. Но лингвистические трансформации не принадлежат к этому уровню, потому что они постулируются, а не выводятся путем индукции.

Итак, в связи с необходимостью преодолеть фундаментальные антиномии в фонологии и грамматике мы пришли к различению двух ступеней абстракции в структурной лингвистике и к соответствующему расщеплению лингвистических понятий, которое приводит нас к построению более сложной концептуальной системы.

В связи с предлагаемым расщеплением лингвистических понятий мы считаем целесообразным ввести в структурную лингвистику понятия лингвистических фенотипов и генотипов. Лингвистические фенотипы — это фонемоиды, дифферентоиды, морфемоиды, синтагмоиды. Лингвистические генотипы — это фонемы, дифференторы, морфемы, синтагмы.

Термины фенотип и генотип заимствованы нами из биологии. «Внешний вид особи в отношении какого-либо признака называется ее фенотипом, а генетическая конституция организма — ее генотипом» [7]. Генотип выражается в символах, как например *BB*, *Bb*, *bb*. Организмы могут иметь одинаковые фенотипы, но разные генотипы, и обратно — одинаковые генотипы, но разные фенотипы.

Мы полагаем, что существует заслуживающая серьезного внимания аналогия между биологическими и лингвистическими фенотипами и генотипами. Так, возвращаясь к нашим примерам, мы видим, что фенотипически сходные фразы *Пение артиста* (фенотип $a \rightarrow b$) и *Чтение книги* (фенотип $a \rightarrow b$) имеют разные генотипы. Генотип первой фразы: Rd ("a" $\leftrightarrow$ "b"). Генотип второй фразы. Rd ["c" $\leftrightarrow$ ("b" $\leftarrow$ "a")]. Обратный случай имеем при сопоставлении безличных и двусоставных предложений. Фенотипически безличные предложения резко отличаются от двусоставных. Но в генотипическом плане безличные предложения тождественны двусоставным предложением. Те и другие имеют одинаковый генотип: "a" $\leftrightarrow$ "b".

Аналогичные факты существуют и в фонологии. Например, если даны две позиции P_1 и P_2 и в обеих позициях встречаются фенотипически сходные звуки типа k, то эти звуки могут быть субстратами разных фонем, то есть они могут иметь разные генотипы. Обратно, звук k в позиции P_1 и звук x в P_2, фенотипически различные, могут быть субстратами одинаковых фонем, то есть они могут иметь одинаковый генотип [8].

В связи с различением лингвистических фенотипов и генотипов целесообразно назвать ступень наблюдения фенотипической ступенью, а ступень конструктов — генотипической ступенью.

PROCESSING OF INFORMATION IN COGNITION AND THE TWO LEVEL THEORY OF STRUCTURAL LINGUISTICS

Summary

The article deals with the principal features of the process of cognition from the view-point of the black box theory in cybernetics. On the basis of this theory and an analogical treatment of cognition process in the modern logic of science it is recommended to distinguish strictly and never confuse two abstraction levels in structural linguistics — the observation level and the level of constructs. The necessity of overcoming the linguistic antinomies considered in the article inevitably leads to the necessity of distinguishing strictly between these abstraction levels. In connection with distinguishing between the observation and the construct level the concepts «linguistic phenotype» and «linguistic genotype» are introduced.

[7] К. Виллиг. Биология. Пер. с англ. М., 1959, стр. 597.

[8] С. К. Шаумян. Двухступенчатая теория фонемы и дифференциальных элементов. «Вопросы языкознания», 1960, № 5, стр. 27.

И. И. Ревзин

О НЕКОТОРЫХ ВОПРОСАХ ДИСТРИБУТИВНОГО АНАЛИЗА И ЕГО ДАЛЬНЕЙШЕЙ ФОРМАЛИЗАЦИИ

(применительно к синтаксическому анализу)

§ 1. В связи с потребностями новых технических приложений науки о языке возникает новая отрасль языковедения — математическая лингвистика, объектом которой являются математические модели тех или иных сторон языка. Всякая модель строится не на голом месте, ей предшествует не только установление некоторых фактов, но также и определенный аппарат лингвистических методов и определенные представления о сущности языка. Этот вопрос уже был предметом обсуждения, причем некоторые авторы считали необходимым «решительно отвергнуть утверждения, будто структурализм является обязательным путем к построению математической лингвистики» [1].

В данной работе мы не ставим своей задачей рассматривать этот спор с общеметодологической точки зрения. Мы хотим лишь показать, что в основу некоторых, наиболее важных, как нам кажется, моделей, рассматриваемых в математической лингвистике, положены именно идеи структурализма и что дальнейшее совершенствование этих моделей целесообразно вести именно с учетом достижений современного структурализма.

Известны далеко идущие аналогии между теоретико-информационным подходом к языку как коду и идеями де Соссюра, которые легли в основу современного структурализма [2] и в той или иной форме принимаются в качестве основных постулатов как представителями европейского структурализма (пражский кружок и школа датского структурализма), так и представителями американской дескриптивной лингвистики. Мы, однако остановимся лишь на одном методе, применяемом в первую очередь в американской дескриптивной лингвистике, именно на методе дистрибутивного анализа, и постараемся показать сходство его исходных предпосылок с аксиоматикой теоретико-множественной модели языка, развитой советскими учеными [3], некоторую

---

[1] Н. Д. Андреев, Л. Р. Зиндер. Основные проблемы прикладной лингвистики. «Вопросы языкознания», 1959, № 4, стр. 19.

[2] См.: B. Mandelbrot. Structure formelle des textes et communication. «Word», vol. X, № 1, April, 1954; его же. An informational theory of the statistical structure of language. «Communication theory». London, 1953. О соотношении между кодом и сообщением в языке см.: Вяч. Вс. Иванов. Код и сообщение. «Бюллетень объединения по проблемам машинного перевода». М., 1957, № 5.

[3] Об основных положениях этой концепции см.: О. С. Кулагина. Об одном способе определения грамматических понятий на базе теории множеств. Сб. «Проблемы кибернетики», № 1, М., 1958, стр. 203—214; И. И. Ревзин. Модели языка. М., 1962.

аналогию в полученных результатах и, наконец, возможность формализации в терминах этой теории некоторых важных понятий структурного анализа. Мы хотели бы также обратить внимание на связь между дистрибуционным анализом и трансформационным анализом, приобретающим все большее значение.

§ 2. Исходное понятие в дистрибуционном анализе — это понятие «совокупности высказываний (utterances), которые могут быть сделаны в данной языковой группе» [4]. Это понятие, в сущности соответствующее речи в понимании Соссюра, и является исходной объективной данностью, единственной реальностью, с которой имеет дело лингвист. Харрис в своем фундаментальном труде, посвященном методам структурного анализа, использует для этого понятия термин universe of discourse [5] (универсальное речевое множество). Перед лингвистом при дистрибуционном анализе стоят три задачи:

1) выделение мельчайших единиц на данном уровне исследования (фонетическом, морфологическом или синтаксическом) или, как принято говорить в дистрибуционном анализе, «сегментация» потока речи;

2) объединение выделенных отрезков в определенные классы (фонемы, морфемы и т. п.);

3) выяснение отношений между выделенными классами на данном уровне исследования.

При решении всех этих задач лингвист и должен пользоваться критерием «дистрибуции» (распределения), т. е. для каждого элемента, предварительно выделяемого на данном уровне, требуется проанализировать все возможные окружения этого элемента в потоке речи. В дальнейшем каждые два элемента сравниваются с точки зрения их возможных окружений и если эти окружения в достаточно большом числе случаев совпадают, то соответствующие элементы объединяются в один класс. Вся эта процедура хорошо описана в цитированной книге Харриса.

Преимущества такой процедуры ясны. Во-первых, при анализе исключается субъективизм, наблюдаемый при анализе, исходящем из значения, т. е. опирающемся на лингвистическую интуицию. Во-вторых, понятие распределения является наиболее общим понятием, с которым мы сталкиваемся в лингвистическом анализе, и может быть применено к анализу л ю б о г о я з ы к а, вне зависимости от особенностей его строя. В-третьих, описанная процедура применима ко всем ярусам языка, и тем самым достигается однообразие лингвистических методов. В-четвертых, все определения в этом анализе суть определения операционные или конструктивные.

Недостатком чисто дистрибуционного анализа является его громоздкость и фактическая трудность его выполнения. На практике многие процедуры сокращаются посредством обращения к значению путем опроса носителей языка (informant). По этому поводу Харрис говорит: «Ситуация, при которой нужно иметь носителя языка, отвечающего на вопросы лингвиста или диктующего ему тексты, не является идеальным источником, но она может оказаться неизбежной при многих лингвистических исследованиях» [6].

Обращение к носителям языка практически полезно для решения следующих двух задач: а) сегментации речевого потока на фонетическом уровне и б) устранения индивидуальных отклонений от нормы. Некоторые ученые считают что первая задача может быть решена спектральным анализом речевого потока [7]. Предположим, что мы решили

---

[4] L. B l o o m f i e l d. A set of postulates. «Language», vol. XI, № 3, 1926, стр. 155.

[5] Z. H a r r i s. Methods in Structure linguistics. Chicago, 1951, стр. 7—9.

[6] Z. H a r r i s. Указ. соч., стр. 14.

[7] См. И. И. Р е в з и н. Об одном подходе к моделям дистрибутивного фонологического анализа (наст. сб., стр. 81).

и вторую задачу, а также еще до применения дистрибутивных методов разбили речь на определенные отрезки — морфемы или слова. «Слово» в дескриптивной лингвистике есть понятие, определяемое на основе понятия «морфема». Соответствующий переход от морфем к словам достаточно сложен, но основная идея сводится к следующему. Некоторые морфемы или сочетания морфем называются «свободными формами», если они не зависят от характера предшествующих морфем, т. е. обладают достаточно свободным распределением; например, нем. *gut arbeiten* (соответственно русск. *хорошо работать*) — свободная форма, а *-en* в *arbeiten* (или *-ть* в *работать*) — связанная форма. Словом тогда можно назвать «минимальную свободную форму». Поскольку, однако, нас интересует в дальнейшем лишь уровень синтаксического анализа, то мы будем считать, что сегментация на слова уже дана. (Для упрощения можно даже считать, что имеется в виду «графическое слово», т. е. цепочка букв от пробела до пробела). Такова, в сущности, исходная ситуация, с которой мы имеем дело в дистрибутивном анализе, когда речь идет о переходе к синтаксическому анализу.

Но мы как раз получили ту ситуацию, которая является исходной в теоретико-множественной концепции языка. В самом деле, там считается заданной некоторая совокупность фраз (ср. вышеприведенное определение Блумфильда). Каждая фраза состоит из слов. Наряду с заданными фразами, которые считаются «отмеченными», позволяется строить и другие последовательности слов, которые, однако, считаются неотмеченными. (В дальнейшем мы будем ссылаться только на ту часть этой концепции, которая не рассматривает никаких других исходных понятий). На этом, однако, сходство обеих концепций не останавливается. В дистрибутивном анализе два элемента, имеющие более или менее одинаковое распределение, объединяются в один класс, в теоретико-множественной концепции два слова x и y считаются эквивалентными, если для любых фраз A и B все фразы вида AxB, AyB являются одновременно отмеченными или одновременно неотмеченными. Совокупность эквивалентных элементов называется «семейством» (например, к одному семейству относятся слова: *корову*, *куклу*, *галошу* и вообще все слова женского рода в винительном падеже ед. числа).

§ 3. Важно отметить и некоторые внутренние отличия системы дескриптивной лингвистики от аппарата определений теоретико-множественной концепции. Мы видели, что при дистрибутивном анализе для эквивалентности двух элементов требуется, чтобы они имели одинаковое окружение в достаточно большом числе случаев, но не обязательно во всех случаях (ср. указание Харриса на практическую невозможность проверки всех окружений) [8]. Теоретико-множественная концепция О. С. Кулагиной есть некоторая м о д е л ь, т. е. идеализация реально существующей ситуации. Необходимость проверки в с е х окружений связана с тем, что эквивалентность в этой модели должна обладать свойствами рефлексивности, симметрии и транзитивности. Как легко проверить, взаимозаменяемость не во всех, а только в достаточно большом числе случаев не обладает свойством транзитивности и поэтому такое понятие не является продуктивным для теоретико-множественной, т. е. л о г и ч е с к о й модели. Другое дело вероятностная модель (например, очень интересная модель Н. Д. Андреева [9], в которой может быть развито это положение дистрибуционного анализа).

Возникает, однако, следующая чисто логическая трудность, связанная с таким методом установления эквивалентности. Для того, чтобы эффек-

[8] Z. H a r r i s. Указ. соч., стр. 244.

[9] Н. Д. А н д р е е в. Моделирование языка на базе его статистической и теоретико-множественной структуры. «Тезисы совещания по математической лингвистике». М., 1959, стр. 15—22.

тивно использовать понятие эквивалентности, множество отмеченных фраз должно быть конечным (например, совокупность всех фраз данной книги, или всех книг данного писателя или даже множество всех книг польских писателей XVI в.). Однако рассмотрение всего множества языка как конечного противоречит открытому характеру реальных языков. Речь (в отличие от языка как системы) должна рассматриваться как бесконечное множество элементов любого уровня [10]. В концепции О. С. Кулагиной и не требуется, чтобы множество отмеченных фраз было конечным. Однако рассмотрение множества отмеченных фраз как бесконечного вызывает трудности при использовании понятия эквивалентности. Чтобы убедиться в эквивалентности двух слов a и b, мы должны перебрать все отмеченные фразы, и этот процесс может длиться бесконечно. На самом деле в концепции О. С. Кулагиной неявно используется абстракция актуальной бесконечности, т. е. сразу берется все (вообще говоря, бесконечное) множество фраз, содержащих a и b. Возможность применения такой абстракции в лингвистике вряд ли может быть доказана. Но как видоизменить понятие эквивалентности таким образом, чтобы не лишать его описанных выше свойств?

Выход легко найти, если обратиться к практике дистрибуционного анализа. Рассматривая возможности упрощения дистрибуционного анализа на морфемном уровне, Харрис говорит: «Метод апроксимации, наиболее широко применяемый в современной лингвистике, состоит в рассмотрении окружений, более коротких, чем полное высказывание. Выбирается ограниченный отрезок каждого высказывания, и морфемы объединяются в класс, если они взаимозаменяемы в этом ограниченном окружении. Так если мы выберем позицию перед морфемой *-ly*, то в один класс объединятся морфемы *large*, *clean*, *true* и т. д.» [11].

В терминах абстрактной теоретико-множественной концепции эта процедура может быть формализована следующим способом. Два слова x и y эквивалентны, если для любых фраз A и B все фразы AxB и AyB, состоящие не более чем из k слов, одновременно отмечены или одновременно не отмечены. Легко проверить, что такая эквивалентность обладает свойствами рефлексивности, симметрии и транзитивности, а с другой стороны, вне зависимости от характера исходного множества фраз проверке каждый раз подлежит лишь конечное множество фраз, длина которых не превышает k слов.

Итак, обращение к практике дистрибутивного анализа помогает улучшить некоторые принципиальные определения теоретико-множественной модели. Это тем более важно, что в теоретико-множественной модели получены некоторые весьма важные результаты, позволяющие, например, формальным путем описывать систему частей речи данного конкретного языка. Но мы не можем останавливаться на вопросах морфологического анализа, так как наша задача — анализ синтаксических проблем. В дальнейшем мы покажем, что некоторые важные синтаксические понятия дескриптивной лингвистики могут быть использованы в теоретико-множественной модели.

§ 4. В структурной дескриптивной лингвистике основным синтаксическим понятием является понятие «непосредственно-составляющие» [12] (в дальнейшем сокращенно: НС). Существуют разные способы введения этого понятия. Мы предпочитаем следующий. Назовем к о н с т р у к ц и е й любую фразу и любую часть фразы. (Обычно под конструкцией

---

[10] Ср.: Н. С. Т р у б е ц к о й. Основы фонологии. Пер. с нем. М., 1960, стр. 8.
[11] Z. H a r r i s. Указ. соч., стр. 255.
[12] См.: L. B l o o m f i e l d. Language. New York, 1933, стр. 160—161; K. L. P i k e. Taxemes and Immediate Constituents. «Language», vol. XIX, № 2, 1943, стр. 70; R. W e l l s. Immediate Constituents. «Language», vol. XXIII, № 2, 1947, стр. 81—117.

понимают только значимую часть фразы [13], но мы хотим оставаться в пределах чисто дистрибуционных терминов; кроме того, «конструкция» есть вспомогательное понятие, которое не обязано соответствовать определенному лингвистическому объекту.) Теперь определим понятие «составляющая данной конструкции» (сокращенно: С). Мы предпочитаем дать рекурсивное определение этого понятия, что соответствует стремлению сделать все лингвистические определения операционными.

Базис этого определения — следующий: 1) Всякое отдельное слово есть С (напомним, что мы для простоты предположили, что каждая фраза уже разбита на слова).

Основная часть определения гласит: 2) Каждая конструкция, имеющая более или менее одинаковое распределение с некоторым С, есть в свою очередь С.

Разберем с этой точки зрения фразу, которую так охотно анализируют грамматисты:

Большая ворона взлетела на высокий куст.

(1) (2) (3) (4) (5) (6).

По пункту 1 нашего определения имеем 6 С:

(1) (2) (3) (4) (5) (6).

По пункту 2 нашего определения находим дополнительно следующие С:

(1) (2) встречаются в тех же окружениях, что и (2);
(5) (6) — в тех же окружениях, что и (6);
(4) (6) — в тех же окружениях, что и слово *вверх*;
(4) (5) (6) — в тех же окружениях, что и (4) (6);
(3) (4) (5) (6) — в тех же окружениях, что (3).

С другой стороны, такие конструкции как *взлетела на* или *на высокий* составляющими не являются, так как нельзя найти составляющую (например, слово), имеющую более или менее одинаковое распределение с этими конструкциями. Конструкции, соответствующие некоторым составляющим, мы будем называть п р а в и л ь н ы м и.

Непосредственно составляющими (НС) данной правильной конструкции назовем С, входящие в данную конструкцию, но не входящие в качестве части в другие С данной фразы. Для дальнейшего важно следующее определение. Если одна и только одна НС в конструкции АВ (например В) имеет то же распределение, что и вся конструкция, то синтаксическое отношение между НС называется подчинением (В подчиняет А). В называется ядром (nucleus), а А спутником (satellite), или атрибутом к В, например: *деревянная* (атрибут) *изба* (ядро); *очень* (атрибут) *приятно* (ядро), *гулять* (ядро) *в парке* (атрибут).

Подчинение может быть многоступенчатым. Например, в конструкции *Цветы очень красивой розы* на первом этапе выделяются НС *цветы* и *очень красивой розы*, где вторая НС подчинена первой, на втором этапе выделяются НС *очень красивой* и *розы*, где первая НС подчинена второй, и т. п.

Здесь оказывается чрезвычайно полезным понятие ранга, введенное еще Есперсеном [14]. Это понятие, как и понятие составляющего, мы определим индуктивно: если некоторая НС является ядром, не подчиненным никакому другому ядру, то она называется НС нулевого ранга. НС, подчиненная ядру, являющемуся НС n-го ранга, называется НС (n+1)-го ранга.

В нашем примере *цветы* есть НС нулевого ранга, *розы* — НС 1-го ранга, *красивой* — НС 2-го ранга, *очень* — НС 3-го ранга.

Несколько по-другому подходит к понятию ранга Харрис [15].

---

[13] См.: Г. Г л и с с о н. Введение в дескриптивную лингвистику. Пер. с англ. М., 1959, стр. 190 и сл.

[14] См.: О. Е с п е р с е н. Философия грамматики, гл. VII. Пер. с англ. М., 1958.

[15] Z. H a r r i s. Указ. соч., стр. 265 и сл.

www.ruslang.ru

Если в предыдущем изложении ранг присваивался атрибутам, т. е. подчиненным элементам, то Харрис предлагает присваивать ранг слову (Харрис, правда, предпочитает говорить о морфемах), имеющему то же распределение, что и некоторая конструкция (например, каждому ядру), в зависимости от характера данной конструкции. Это необходимо для уточнения одинаковости распределения. В самом деле, мы утверждаем, что *красивой розы* имеет то же распределение, что и *розы*. Это верно только в том случае, если не рассматривать сочетаний типа *очень красивой розы* (здесь замену *красивой розы* на *розы* произвести нельзя из-за недопустимости сочетания *очень розы*). Если же считать, что слово *розы* в самостоятельном употреблении имеет 2-й ранг, именно потому, что взаимозаменяемо с конструкцией *очень красивой розы*, где у слова *розы* «нулевой ранг», то мы не получим никаких противоречий. Если слово имеет более высокий ранг в самостоятельном употреблении, то это значит, что оно имеет более низкий ранг в некоторых конструкциях, в то время как обратное неверно. Мы далее увидим, сколь близки эти идеи Харриса к пониманию ранга в теоретико-множественной концепции.

§ 5. Основным синтаксическим понятием теоретико-множественной концепции является понятие конфигурации. Вот определение этого понятия (для простоты, как и при рассмотрении проблемы эквивалентности, рассматриваются цепочки слов, а не классов слов):

1) Любая фраза считается фразой нулевого ранга. Некоторая часть фразы T_1, состоящая из двух или более слов, называется конфигурацией 1-го ранга, если найдется слово x, такое, что при любых окружениях A и B все фразы нулевого ранга вида AT_1B и AxB одновременно отмечены или не отмечены. Слово x называется результирующим конфигурации 1-го ранга. Фраза, не содержащая конфигураций 1-го ранга, называется фразой 1-го ранга.

2) Некоторая часть фразы T_n, состоящая из двух или более слов, называется конфигурацией n-го ранга, если найдется слово x, такое, что при любых окружениях A и B все фразы (n+1)-го ранга вида AT_nB и AxB одновременно отмечены или одновременно не отмечены. Слово называется результирующим конфигурации n-го ранга. Фраза, не содержащая конфигураций n-го ранга, называется фразой n-го ранга.

Легко видеть, что этим определением все множество фраз разбивается на подмножества таким образом, что все фразы n-го ранга одновременно являются фразами (n—1)-го ранга, в то время как обратное неверно. Поэтому любое множество фраз n-го ранга как бы вложено во множество фраз (n—1)-го ранга и все множества фраз любого ранга как бы вложены во множество фраз нулевого ранга.

Особенности этого определения в следующем:

а) оно носит рекурсивный характер, как и приведенные выше определения составляющих и рангов;

б) два понятия: «конфигурация n-го ранга» и «фраза n-го ранга» определяются одновременно. Определения такого вида, когда одновременно определяется несколько понятий, вполне допустимы в современной логике[16];

в) каждый ранг фразы включает в себя все предыдущие, т. е. каждая фраза n-го ранга является одновременно фразой (n—1)-го ранга. Подобная ситуация с рангами ядер (см. § 4) описана Харрисом [17];

г) определение конфигурации, так же как и определение § 2, вводит идею эквивалентности со свойствами симметрии, рефлексивности и транзитивности, однако здесь эквивалентность каждый раз определяется на некотором ограниченном множестве фраз, а именно на множестве

---

[16] См.: Н. А. Ш а н и н. О некоторых логических проблемах арифметики. «Труды Математического ин-та им. В. А. Стеклова», т. 43, 1955, стр. 3—5.

[17] Z. H a r r i s. Указ. соч.

фраз соответствующего ранга. Ниже мы увидим, что и здесь, как и в § 2, целесообразно ограничить длину рассматриваемых окружений.

Все эти особенности данного определения делают его чрезвычайно мощным орудием для построения абстрактного синтаксиса в терминах рассматриваемой модели. Мы не можем останавливаться на достигнутых результатах. Они даны в «Моделях языка». Отметим лишь, что в терминах конфигурационного анализа могут быть формализованы понятия, рассмотренные в § 4. Что такая формализация возможна, следует прежде всего из общности исходных предпосылок и общности метода. Покажем теперь, что на основе понятия конфигурации, которое, как легко увидеть, соответствует понятию правильной конструкции, может быть формализовано понятие НС.

Прежде всего введем понятие регулярной конфигурации. Мы будем называть конфигурацию регулярной, если ее результирующий совпадает с самым правым словом конфигурации, или с самым левым словом конфигурации. Результирующий регулярной конфигурации n-го ранга называется ядром n-го ранга, а остальная часть регулярной конфигурации атрибутом n-го ранга.

Легко проверить, что эти понятия целиком соответствуют введенным ранее, и все дальнейшее автоматически переносится в теорию конфигурационного анализа. Любопытно, однако, следующее. Если не ограничивать длину проверяемых окружений (как мы, следуя О. С. Кулагиной, это сделали в определении конфигурации), то понятие регулярной конфигурации становится практически бесполезным. А именно в практических приложениях (в отличие от общей теории) мы всегда имеем дело с конечным множеством отмеченных фраз.

Между тем справедливо следующее утверждение. На конечном множестве отмеченных фраз не может быть выделена ни одна регулярная конфигурация.

Исходя из этих соображений, а также из общих соображений, высказанных в § 2, следует и здесь ограничить длину рассматриваемых окружений. В этом случае ничто не препятствует выделению регулярных конфигураций на конечном множестве фраз.

§ 6. Сопоставляя понятия дескриптивной лингвистики и теоретико-множественной концепции, мы неоднократно устанавливали сходство рассматриваемых понятий. При таком сопоставлении может создаться впечатление, что теоретико-множественная модель пересказывает в других терминах все то, что уже было сказано в дескриптивной лингвистике. Это далеко не так. Прежде всего следует отметить, что идеи дескриптивной лингвистики в данной работе освещены, так сказать, сквозь призму теоретико-множественной модели. Как уже приходилось указывать автору, неразработанность методики и даже терминологии дескриптивной лингвистики, громоздкость ее аппарата и неясность исходных предпосылок делает изучение дескриптивных методов делом весьма трудным и мало продуктивным [18].

Нельзя сказать, чтобы идея моделирования была полностью чужда дескриптивной лингвистике. Так, Харрис сочувственно цитирует следующие слова Карнапа: «Непосредственный анализ языка не может не окончиться неудачей, так же как и физик был бы беспомощен, если бы он с самого начала попытался соотносить свои законы с конкретными объектами-деревьями и т. п. Вместо этого он соотносит свои законы с простейшими конструктами: идеальными рычагами, точечными массами и т. д.» [19].

---

[18] См., например: О соотношении структурных и статистических методов в современной лингвистике. «Вопросы статистики речи» (Материалы совещания). Л., 1958, стр. 51.

[19] R. Carnap. Logical Syntax of Language. New York, 1938, стр. 8.

Далее Харрис указывает, что в то время как логики, избегая описанной Карнапом трудности, строят искусственные языки, дескриптивная лингвистика изучает конкретные языки, «но вместо того, чтобы брать части действительных речевых данностей (occurences) в качестве элементов, лингвисты конструируют весьма простые элементы, которые только соотносятся (are associated) с соответствующими чертами речевых данностей» [20].

Однако эта идея проводится далеко не последовательно, и именно простоты, которую декларирует Харрис, в дескриптивной лингвистике достичь не удалось.

Только создание подлинных моделей языка с четкой аксиоматикой и строгим аппаратом определений помогает действительно формализовать ситуацию, рассматриваемую в дескриптивной лингвистике. Важно, однако, подчеркнуть, что именно сходство исходной ситуации в дескриптивной лингвистике и теоретико-множественной модели позволяет обогатить эту модель обращением к идеям дескриптивной лингвистики. Таким несомненным обогащением теоретико-множественной модели и является, с нашей точки зрения, анализ по непосредственно составляющим.

§ 7. Введение понятия НС в теоретико-множественную модель важно и из следующих соображений. Как с точки зрения интересов практических приложений (например, машинного перевода), так и с точки зрения теории целесообразно для всякой модели анализа некоторого множества фраз иметь соответствующую модель порождения данного множества фраз из некоторых исходных данных [21]. Такой обратной моделью по отношению к модели О. С. Кулагиной является модель порождения методом НС, рассмотренная Н. Хомским [22]. В этой модели задается система правил, позволяющая путем последовательного разложения на НС получить некоторое множество отмеченных фраз. Интересно, что можно построить и более сильную модель порождения, а именно модель трансформационную, наиболее простым частным случаем которой и является модель порождения по НС. Эта новая модель позволяет описать ряд фактов, которые не могут быть описаны методом порождения по НС. Например, методом порождения по НС нельзя получить такие фразы, как русск. *Сегодня я болен* или нем. *Gestern baute er das Haus.* Дело в том, что эти фразы, как легко проверить, не разлагаются линейно на НС. Можно показать, что применение к этим фразам аппарата определений теоретико-множественной модели не дает результатов, адекватных нашему интуитивному представлению. В частности, *сегодня* окажется атрибутом к *Я*, а приведенная немецкая фраза уже не будет содержать никаких других конфигураций.

Возникает вопрос, можно ли дополнить теоретико-множественную модель такими определениями, чтобы, исходя из дистрибуции и только из дистрибуции, получить результаты, адекватные нашей грамматической интуиции. По крайней мере для решения проблем, возникающих в связи с отсутствием непрерывности в расположении составляющих, можно ввести понятие т р а н с п о з и ц и и, целиком основанное на дистрибуции.

Определение. Простой транспозицией (или транспозицией характера два) фразы n-го ранга мы будем называть такую перестановку двух рядом стоящих слов x и y, при которой фразы n-го ранга (длины не больше k) A xy B и A yx B одновременно отмечены или неотмечены для любых A и B. Аналогично определяется транспозиция характера l, где l число переставляемых слов.

---

[20] Z. H a r r i s. Указ. соч., стр. 15.

[21] Ср.: С. Л. Ф и т и а л о в. Формально-математические модели языков и структура алгоритмов перевода. «Тезисы совещания по математической лингвистике». Л., 1959, стр. 44.

[22] N. C h o m s k y. Syntactic Structures. 's-Gravenhage, 1957.

Теперь, произведя транспозицию характера два в приведенном русском предложении, мы приведем его к виду *Я сегодня болен*, где *сегодня болен* есть регулярная конфигурация с ядром *болен* и атрибутом *сегодня*. В приведенном немецком предложении любая транспозиция характера два приведет, как легко проверить, к неотмеченной фразе. Но транспозиция характера 3 дает фразу *Er baute gestern das Haus*, где последовательно выделяются конфигурации *baute gestern*, *baute das Haus*.

Как мы видим, дистрибутивные критерии оказываются полезными и здесь. Остается, однако, еще нерешенной очень важная проблема: в какой мере выполнима общая задача построения теоретико-множественной модели, исходящей из дистрибуции, такой, чтобы она целиком соответствовала трансформационной модели порождения, построенной Хомским. Решение этой задачи зависит, по-видимому, и от четкой логической формулировки правил трансформационного анализа.

ON SOME QUESTIONS OF DISTRIBUTIONAL ANALYSIS AND ITS FURTHER FORMALIZATION

Summary

There are some remarkable links between distributional analysis as applied by American descriptive linguists and the point of departure of the so-called set-theoretical conception of language developed by Soviet scholars.

One can hope therefore, that many concetps applied by descriptivists and some solutions proposed by them would enlarge the possibilities and linguistic applicability of set-theoretical models of language, would make them more revealing from the logical point of view.

The author shows further how such fundamental concepts of descriptive analysis as that of «immediate constituents» can be formalized in set-theoretical terms.

In conclusion, the author makes some remarks on the relation between generative grammars as proposed by Chomsky, and distributional grammars. In order to connect ideas of transformational and distributional analysis, the concept of «transposition» is introduced.

С. Маркус

ТЕОРИЯ ГРАФОВ, ЛИНГВИСТИЧЕСКИЕ ОППОЗИЦИИ И ИНВАРИАНТНАЯ СТРУКТУРА[1]

Во многих исследованиях о форме языка в конечном счете исследуется множество элементов, между которыми существуют те или иные отношения. Среди этих отношений наиболее простыми являются бинарные отношения, то есть отношения, определенные для некоторых пар элементов, составляющих все множество. Вот несколько примеров бинарных отношений, которые более или менее явно появляются в наиболее важных лингвистических исследованиях (обозначим через X совокупность рассматриваемых элементов и через ρ бинарное отношение, определяемое для некоторых пар элементов множества X):

1) X — *множество слов модели* Кулагиной; *для* $x \in X$, $y \in X$ *имеем* $x\rho y$ (т. е. x находится в отношении ρ с y), если x и y принадлежат к одной окрестности [10];

2) X — то же, что и в пункте 1; для $x \in X$, $y \in X$ имеем $x\rho y$, если x и y принадлежат к одному семейству [10];

3) X — то же, что и в пункте 1; язык предполагается простым; ρ — принадлежность к тому же классу [10];

4) X — то же, что и в пункте 1; язык предполагается простым; ρ — принадлежность к тому же типу [10];

5) Y — множество фонем языка, $X = Y \times Y$ ($Y \times Y$ есть произведение Y на самого себя и определяется как множество пар (y, z) (где $y \in Y$, $z \in Y$); для $y \in Y$, $z \in Y$, $y' \in Y$, $z' \in Y$ $x = (y, z)$, $x' = (y', z')$ имеем $x\rho x'$, если оппозиция между y и z имеет ту же основу, что и оппозиция между y' и z' (многосторонние оппозиции Трубецкого) [17];

6) X — то же, что и в пункте 5; $x\rho x'$, если y и z имеют те же дифференциальные черты, что и y' и z' (пропорциональные оппозиции Трубецкого) [17];

7) X — множество звуков языка; $x\rho y$, если x и y никогда не дают фонологической коммутации (замены);

8) X — множество слов; $x\rho y$, если существует хотя бы один контекст, общий для x и y (З. Харрис) [8];

9) X — множество грамматем (грамматемой мы называем множество черт, родственных в морфологическом отношении, например: мужской род, единственное число, положительная степень, родительный падеж); $x\rho y$, если y зависит от x (Мартемьянов, Ревзин); [12], [15];

10) X — то же, что и в пункте 9; $x\rho y$, если x может хотя бы раз доминировать над y (Мартемьянов) [12];

11) Тот же пример, что в пункте 5, но с заменой фонем грамматемами;

---

[1] Автор благодарит А. С. Желябина (Институт языкознания АН СССР) за перевод статьи на русский язык.

12) Тот же пример, что в п. 6, но с заменой фонем грамматемами;

13) X — то же, что и в п. 7; $x\rho y$, если хотя бы раз имеется коммутация между x и y;

14) X — множество фонем или грамматем; $x\rho y$, если оппозиция между x и y хотя бы раз нейтрализуется;

15) X — множество грамматических категорий; $x\rho y$, если присутствие x влечет за собой присутствие y (селективность по Ельмслеву) [9];

16) X — то же, что и в пункте 15; $x\rho y$, если присутствие x влечет за собой присутствие y, и наоборот (солидарность по Ельмслеву) [9];

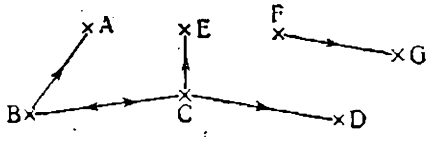
17) X — множество возможных состояний языка с конечным числом возможных состояний (Н. Хомский — Г. Миллер); $x\rho y$, если можно перейти от состояния x к состоянию y [5];

18) X — то же, что и в п. 14; $x\rho y$, если оппозиция между x и y отрицательна (привативна) в пользу x (Трубецкой) [17];

19) X — множество контекстов; $x\rho y$, если x и y могут быть контекстами для одного и того же слова (синтагматические оппозиции Ф. де Соссюра);

20) X — то же, что и в п. 1; $x\rho y$, если в каждой фразе, содержащей x, замена x на y приводит к специальной фразе (Р. Л. Добрушин) [6].

Все эти примеры входят в одну и ту же схему. Рассматривается множество X элементов безразличной природы и закон ρ, который ставит в соответствие к каждому элементу x системы X некоторое подмножество $\rho(x)$ элементов X. (Это подмножество образуется в названных выше примерах из элементов, с которыми x находился в отношении ρ). Множество X и закон ρ образуют то, что называется в математике г р а ф. Граф изображается геометрически совокупностью точек и совокупностью дуг, связывающих некоторые из этих точек:



Точки A, B, C, D, E, F, G представляют элементы множества X; наличие дуги, например, той, которая идет от B к A, указывает, что $B\rho A$.

Для каждого графа существует также м а т р и ч н о е п р е д с т а в л е н и е, т. е. представление в виде таблицы, в которой каждому элементу графа соответствует определенная строка и определенный столбец. На пересечении строки, соответствующей X, со столбцом, соответствующим Y, ставят 1, если $X\rho Y$, и ставят 0 в обратном случае. Например, для графа, изображенного геометрически выше, будем иметь следующее матричное представление:

	A	B	C	D	E	F	G
A	0	0	0	0	0	0	0
B	1	0	1	0	0	0	0
C	0	1	0	1	1	0	0
D	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0
F	0	0	0	0	0	0	1
G	0	0	0	0	0	0	0

Теория графов получила в математике последних лет огромное развитие, что в некоторой мере объясняется ее применением в автоматике. Хотя эта теория еще не была развита в применении к лингвистике, многие из понятий и результатов теории графов относятся к разнообразнейшим сторонам языковых явлений. В лингвистике могут быть использованы не

только основные понятия теории графов, но также некоторые очень глубокие результаты. При этом лингвистика ставит перед теорией графов новые проблемы, которые должны учитываться математиками.

Легко проверить, что в рассмотренных выше примерах 2, 3, 4, 5, 6, 11, 12 и 16 отношение ρ обладает следующими свойствами: для каждого $x \in X$ имеется $x\rho x$ (свойство рефлексивности); если $x \in X$, $y \in X$ и $x\rho y$, то $y\rho x$ (свойство симметрии); если $x \in X, y \in X$, $z \in X$, $x\rho y$ и $y\rho z$, то $x\rho z$ (свойство транзитивности). Бинарное отношение, которое обладает такими свойствами, является о т н о ш е н и е м э к в и в а л е н т н о с т и. Отношения эквивалентности приводят к простейшим графам, которые были исследованы до того, как теория графов сложилась как самостоятельная теория. Применение отношений эквивалентности к лингвистике является весьма важным, как явствует из последующих исследований, например, из работ Шаумяна [19], Кулагиной [10], Ревзина [15], Добрушина [6]. С другой стороны, следует отметить, что имеются другие весьма важные ситуации, когда отношения эквивалентности не были выявлены; таковы отношения, рассмотренные в примерах 5 и 6. При этом небходимо подчеркнуть, что теория оппозиций Трубецкого не получила еще полного пересмотра, необходимого для того, чтобы привести ее в соответствие с общностью и точностью, требуемыми современным развитием лингвистики. Правда Ж. Кантино значительно продвинул вопрос, произведя логическую классификацию оппозиций и развив теорию сигнификативных оппозиций [3], [4]; в свою очередь Ж. Апостель и Б. Мандельброт оказали большую услугу науке, разработав аналогии между лингвистической оппозицией и расстоянием Хэмминга, рассматриваемым в теории кодов [1]. Но, как мы увидим, необходимо сделать еще кое-что, и ответ следует искать в теории графов.

Прежде всего, эта теория обязывает нас уточнить круг рассматриваемых элементов, то есть множество X. Таким образом можно избегнуть ситуаций, когда не известно точно, идет ли речь об оппозиции между двусторонними знаками или лишь между элементами в плане выражения.

Для определенности предположим, что X — множество фонем языка. Отношение ρ в этом случае таково: для $x \in X$, $y \in X$ имеем $x\rho y$, если x обладает хотя бы одной чертой, не принадлежащей y. Таким образом, получаем граф, который дает: п р и в а т и в н ы е оппозиции, если имеем $x\rho y$, но не $y\rho x$; э к в и п о л л е н т н ы е оппозиции, если имеем и $x\rho y$, и $y\rho x$; н у л е в ы е оппозиции, если нет ни $x\rho y$, ни $y\rho x$ (когда x совпадает с y).

Различие между этими типами оппозиций было ясно показано Ж. Кантино при помощи отношения включения теории множеств [4]; подобное отношение всегда порождает граф. Но мы покажем, что использование графов ведет к гораздо более общему и более полному описанию типов оппозиций, или, как говорит Ж. Кантино, лингвистических отношений (и не только лингвистических). Действительно, точка зрения Кантино может применяться лишь в случае, если элементы множества X могут рассматриваться как множества более примитивных элементов и отношение ρ является отношением включения. Но в лингвистических исследованиях последнего времени имеются все более и более многочисленные случаи, когда это невозможно; отношение ρ иногда бинарно и не может быть сведено к отношению включения между элементами графа, рассматриваемыми как множества более примитивных элементов; с другой стороны, нельзя отказать таким отношениям в праве быть лингвистическими оппозициями. Таковы случаи в примерах 7, 8, 10, 13, 14, 17; таково также отношение «больше» ($>$), рассматриваемое в модели, предложенной И. И. Ревзиным ([15]. В таких случаях описание Кантино больше не имеет смысла, но можно еще использовать граф, поскольку всегда имеются бинарные отношения. (Следует отметить, что градуальные оппозиции,

сведенные Кантино к оппозициям привативным, могут быть также описаны более общим образом при помощи теории графов.)

Исследовав лингвистическую оппозицию с точки зрения отношения, существующего между ее членами, необходимо описать, следуя Трубецкому, сходства и различия между оппозициями. Для этого введем новый граф, элементами которого будут сами оппозиции, а закон ρ будет избран в зависимости от критерия, по которому мы будем сравнивать, сопоставлять оппозиции. В частном случае, когда элементы исходного графа являются множествами более примитивных элементов (такими множествами являются фонемы и грамматемы), есть две возможности: сравнивать оппозиции с точки зрения их основы, то есть пересечения членов оппозиции, или с точки зрения их отличительных черт. Тогда в случае фонем получим графы, рассмотренные в пунктах 5 и 6, а в случае грамматем — графы, рассмотренные в пунктах 11 и 12. Есть, следовательно, две классификации фонематических и грамматических оппозиций на классы эквивалентности, причем если две оппозиции принадлежат в одной из классификаций к одному классу, то в другой они принадлежат к разным классам. Каждый класс эквивалентности содержит некоторое количество оппозиций, измеряющее вес этих оппозиций или, вернее, их типа в имманентной системе языка; таким образом, возникает нечто вроде иерархии типов оппозиций, и было бы интересно изучить отношение, которое устанавливается между двумя классами эквивалентности, содержащими данную оппозицию. Было бы также еще интереснее рассмотреть для каждого типа оппозиции ее вероятность в языке.

Очевидно, что упомянутые выше два деления оппозиций на классы эквивалентности соответствуют рядам оппозиций (включая корреляции), рассмотренным Трубецким; двусторонние оппозиции и оппозиции изолированные соответствуют классам эквивалентности, состоящим из одной оппозиции. Но здесь следует отметить, что выдвижение на первый план различий «двусторонний — многосторонний» и «изолированный — пропорциональный» довольно искусственно. Ж. Кантино указывает, что сигнификативные оппозиции тогда и только тогда принадлежат к грамматике, когда они пропорциональны. Он утверждает, что оппозиция *sec—sèche*, поскольку она является изолированной, перестает поэтому принадлежать к грамматике [3]. Но тогда получается, что если имелось бы еще одно прилагательное, переход которого из мужского рода в женский осуществляется таким же образом, то можно было бы сказать, что оппозиция *sec—sèche* принадлежит к грамматике. То есть, наличие или отсутствие одного слова решает вопрос о принадлежности или непринадлежности того или иного языкового явления к грамматике. Совершенно ясно, что это нельзя считать хорошим критерием. Главным фактом здесь является классификация оппозиций на классы эквивалентности по многосторонности и пропорциональности, частным случаем чего являются корреляции Трубецкого. Правда, терминология Трубецкого не является адэкватной, но мы не будем здесь обсуждать этот вопрос.

Следует отметить, что описанные выше классы эквивалентности имеют смысл в том случае, когда элементы исходного графа, где оппозиции определены, являются множествами более простых элементов.

Вообще при рассмотрении «лингвистического поля» X, то есть множества элементов, имеющих определенное лингвистическое значение, и при определении в X бинарного отношения ρ получаем граф G; может случиться, что, используя определенное свойство оппозиций, введенных ρ, получим отношение эквивалентности в $X \times X$, а следовательно, граф, элементы которого являются элементами из $X \times X$ и бинарное отношение является именно отношением эквивалентности. Каждый класс эквивалентности определяет вполне связный подграф, то есть подграф, в котором каждая пара элементов соединяется дугой.

Но можно классифицировать оппозиции, взяв в качестве критерия бинарное отношение, которое не является отношением эквивалентности; в этом случае получаем граф, элементы которого также являются элементами из $X \times X$, но природа которого более сложна. Выбор бинарного отношения определяется рассмотренной конкретной проблемой, но небезынтересно рассмотреть также некоторые абстрактные классификации.

Классификация оппозиций на «постоянные» и «нейтрализуемые» находит свое описание в графе типа, рассмотренного в примере 14.

Теория графов дает не только описание лингвистических оппозиций, но также и способы изучения формальных свойств оппозиций. Иногда удается пролить свет, в специфических терминах этой теории, на некоторые особенности инвариантной структуры языка. Для иллюстрации рассмотрим несколько примеров.

Рассмотрим две фонемы, согласную и гласную, и зададимся вопросом, сколько есть возможностей перейти от одной к другой, если предположить, что первая находится в начале инициальной консонантной группы, образованной тремя согласными, а другая расположена непосредственно после группы. Для румынского языка эта проблема решена в [11].

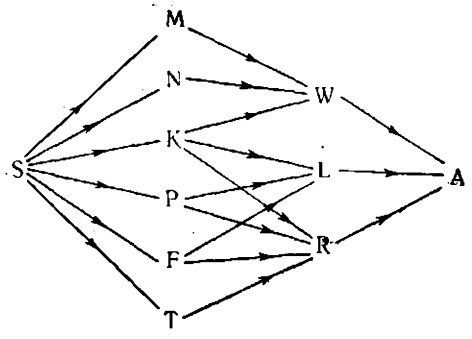


Схема перехода от S к A следующая.

Здесь имеем граф, дугами которого являются стрелки, указывающие возможности перехода. Теория графов дает простое средство определить число γ возможностей перехода от S к A в этом графе, не прибегая к геометрическому представлению графа, а лишь имея в виду дуги и элементы, например, при помощи матричного представления: из числа дуг вычтем число элементов и прибавим 2; тогда получим $\gamma=19-11+2=10$. Разумеется, этот метод более эффективен, когда граф более сложен.

При помощи полученного таким образом числа можно определить энтропию e_0 (нулевого порядка) перехода от S к A в указанных позициях, то есть неопределенность, устраняемую знанием согласных, через которые совершается переход от S к A. Получаем:

$$e_0=\log_2\gamma=\log_2 10.$$

Мы предполагали до сих пор, что вероятность появления каждой фонемы одинакова, и не учитывали, что после определенных фонем одни фонемы появляются чаще, чем другие. Теория графов выработала методы для рассмотрения вероятности. Например, если каждая дуга указанного графа снабжена положительным числом, равным вероятности перехода от одного конца дуги к другому, то в этом случае получим то, что в теории графов называется с е т ь ю.

Теория сетей уже показала свою эффективность в так называемой «задаче о перевозках». Эту теорию можно связать с энтропией второго порядка.

Подобная же задача возникает для грамматем. Согласно аналогии С. К. Шаумяна [18], имеются следующие соответствия: гласная фонема — сказуемое; начальная консонантная группа — группа подлежащего; конечная консонантная группа — группа дополнения.

Для групп подлежащих можно сделать то же самое, что в [11] сделано для начальных консонантных групп; тогда получим графы, элементами

которых являются грамматемы, и две грамматемы соединены дугой, если они могут находиться одна рядом с другой в группе подлежащего. Можно подсчитать число возможностей перехода одной грамматемы в другую.

Но можно рассматривать также фонологическую систему и систему грамматем в целом; в соответствующих графах можно получить сведения о фонологической энтропии и грамматической энтропии языка, рассматриваемого как марковский процесс.

Рассмотрим теперь новый граф, элементами которого являются фонемы или грамматемы и два элемента x и y находятся в отношении $\rho : xpy$, если расположение рядом xy возможно в определенном языке, фиксированном заранее (в настоящей статье примеры будут даваться на материале румынского языка).

Чтобы увидеть, в какой степени используются возможности комбинирования фонем и грамматем, интересно определить максимальное число элементов, которые не могут находиться рядом по два никоим образом. Это число называется числом внутренней устойчивости графа. Любое множество элементов графа, которые не могут находиться рядом попарно, внутренне устойчиво. Для графа, образованного фонемами начальных консонантных групп румынского языка в [11], установлено следующее внутренне устойчивое множество: B, D, G, M, T, V, H, Ţ; следовательно, число внутренней устойчивости выше 7.

Что касается внутренне устойчивых множеств, то в этом отношении не лишена лингвистического значения следующая теорема Ремзея [16]: каждому натуральному числу n соответствует натуральное число $g(n)$, такое, что всякий граф, образованный по крайней мере $g(n)$ элементами, содержит подграф, вполне связанный с n элементами, или внутренне устойчивый подграф с n элементами. Эрдеш установил, что

$$g(n) \leqq \binom{2n-2}{n-1},$$

где выражение $\binom{2n-2}{n-1}$ обозначает число сочетаний $(2n-2)$ предметов, взятых по $(n-1)$ [7]. Если число грамматем выше или равно $\binom{2n-2}{n-1}$, а число внутренней устойчивости графа грамматем ниже n, тогда существует по крайней мере n грамматем, которые могут находиться рядом попарно во всех возможных сочетаниях: обратно, если всякий вполне связный подграф имеет n элементов и если число грамматем не превосходит $\binom{2n-2}{n-1}$, то число внутренней устойчивости больше или равно n.

Множество, состоящее из элементов графа, внешне устойчиво, если от каждого элемента, не принадлежащего этому множеству, отходит дуга, конец которой совпадает с некоторым элементом данного множества.

Для графа начальных консонантных групп румынского языка в [11] установлено, что полусогласные y и w образуют внешне устойчивое множество. Это множество — наименьшее внешне устойчивое множество, и именно потому число его элементов, равное 2, является числом внешней устойчивости графа. Следует отметить, что фонемы y и w не могут находиться рядом ни в одной исходной консонантной группе, и следовательно, они также образуют внутренне устойчивое множество. Такое множество, являющееся внутренне устойчивым и вместе с тем внешне устойчивым, называется в теории графов ядром или базисом. Наличие ядра в графе начальных исходных консонантных групп не должно нас удивлять; в теории графов показано, что каждый граф с конечным числом элементов

обладает ядром. Чтобы лучше подчеркнуть лингвистическое значение ядра, рассмотрим еще два примера.

1) Известно, что в румынском языке очень многочисленны слова, где две рядом стоящие фонемы являются одна согласной, другая гласной. Ограничившись такими словами, рассмотрим граф фонем румынского языка, где имеем xy, если наличие рядом xy возможно по крайней мере в одном слове языка. Поскольку каждая согласная может иметь после себя по крайней мере одну гласную, отсюда следует вывод, что гласные образуют ядро румынского языка. Этот факт, может быть, справедлив с некоторым приближением для многих других языков, поскольку он рассматривается бессознательно при построении стенографии, где, как известно, гласные опускаются именно потому, что они «вставлены» между согласными.

2) рассмотрим упрощенную структуру предложений румынского языка: существительное — прилагательное — глагол — наречие.

Если свести румынский язык к предложениям по указанной схеме, то граф грамматем румынского языка допускает ядро, образованное 41 грамматемой: 40 получается при сочетании различных грамматических значений прилагательного, исключая категории, выражающие степени сравнения (два значения определения, 2 — рода, 2 — числа и 5 падежных; их произведение составляет 40) и одно значение соответствует наречию, взятому в случайной степени сравнения. Следует отметить, что ядро образовано исключительно из грамматем, зависимых справа (по терминологии Мартемьянова), тогда как все слова, которые не принадлежат к ядру, являются главными или базисными (по терминологии того же автора).

Величина ядра дает некоторые сведения о степени использования комбинаторных возможностей фонем или грамматем. Но для того, чтобы эти сведения были инвариантными по отношению к выбору ядра, следует убедиться, что число элементов ядра некоторым образом независимо от выбора ядра. Теория графов дает вполне удовлетворительный ответ на этот вопрос: есть теорема, убеждающая нас в том, что все ядра графа с конечным числом элементов состоят из одного и того же числа элементов. Следовательно, достаточно, чтобы тем или иным, хотя бы эмпирическим путем было найдено ядро, чтобы можно было утверждать, что число элементов этого ядра составляет объективную характеристику, присущую языку. Эта характеристика будет фонологической или грамматической в зависимости от рассматриваемого графа: графа фонем или графа грамматем.

В заключение рассмотрим еще одно лингвистическое приложение теории графов; возможно, этот пример наиболее убедителен. Известно, что языковое сообщение лучше воспринимается, если оппозиция между двумя следующими один за другим элементами этого сообщения резко выражена. С этой точки зрения интересно изучить проблему различения соседних фонем по участию голоса, месту или способу артикуляции. Э. Василиу и я изучили эту проблему для консонантизма румынского языка. Таким же образом можно изучить проблему различения соседних грамматем с точки зрения частей речи, которые они представляют. Вот как ставится и разрешается подобная проблема при помощи теории графов.

Предположим, что мы можем окрасить элементы графа при помощи p цветов таким образом, чтобы два элемента, связанные между собой дугой, не были окрашены в один и тот же цвет. В этом случае можно сказать, что граф P хроматичен. Наименьшее число p, для которого граф P хроматичен, называется хроматическим числом графа. Еще одно определение: если можно окрасить элементы графа так, чтобы два элемента, соединенные дугой, не были одного цвета, можно сказать, что граф хроматичен по отношению к данной окраске.

Изучим теперь различия соседних согласных в начальных консонантных группах с точки зрения локализации. Имеется четыре типа локализации: лабиальная, зубная, палатальная и велярная. Две соседних согласных не всегда различны с точки зрения типа локализации. Естественно возникает вопрос, вызвано ли отсутствие отличия только причинами лингвистического порядка, происходящими исключительно из природы человеческого языка, или оно определяется также нелингвистическими причинами. Теория графов позволяет решить эти вопросы. И действительно, придадим каждому типу локализации определенный цвет: лабиальная — синий, зубная — зеленый, палатальная — красный, велярная — желтый и окрасим каждый элемент графа начальных консонантных групп в цвет, соответствующий его типу локализации. Тот факт, что соседние согласные не всегда отличаются по типу локализации, выражается в нехроматическом характере графа. Теперь вопрос таков: объясняется ли нехроматический характер графа исключительно лингвистическими причинами или же также некоторыми нелингвистическими причинами? Если можно показать, что хроматическое число графа больше четырех (то есть больше количества цветов), справедлив второй вариант альтернативы; следовательно, независимо от всякой лингвистической причины или причины, связанной с природой человеческого языка, дифференциация по локализации в с е х соседних согласных в начальных консонантных группах румынского языка невозможна. Иными словами, если хроматическое число выше четырех, то не нужно уточнять, что речь идет о согласных румынского языка; достаточно сказать, что имеется 20 элементов, и назвать 79 пар этих элементов, связанных дугами, чтобы иметь возможность заключить, что дифференциация находящихся рядом элементов по критерию, принимающему лишь три различных значения, невозможна. Именно таково реальное положение, так как в [11] показано, что хроматическое число рассмотренного графа выше четырех.

Аналогичные задачи решаются и в отношении дифференциации по участию голоса и способу артикуляции, но мы больше не останавливаемся на таких примерах. Таким же образом можно рассматривать вопрос дифференциации соседних грамматем в субъектных или дополнительных группах или вообще в языке, взяв, например, в качестве критерия различения часть речи, соответствующую грамматеме. Придадим каждой части речи определенный цвет и изучим хроматическое число графа. Ясно, что этот граф не хроматичен по отношению к рассматриваемой окраске, поскольку можно иметь две рядом стоящие грамматемы, которые и представляют одну и ту же часть речи. Если хроматическое число больше числа частей речи, можно сказать, что возможность поставить рядом две грамматемы, принадлежащие к одной и той же части речи, объясняется не только лингвистическими причинами, но также некоторыми нелингвистическими причинами, носящими комбинаторный характер.

Следует обратить внимание на задачу, которая совершенно естественно теперь возникает; при наличии графа и его определенной окраски определить минимальное число элементов, которые нужно удалить из графа, чтобы оставшийся граф был хроматичным. Э. Василиу и я решили эту проблему для некоторых окрасок лингвистического значения графа начальных консонантных групп румынского языка. Мы установили, например, что необходимо и достаточно удалить из этого графа пять элементов, избранных соответствующим образом, чтобы оставшийся граф был хроматическим, если окраска дана по типам локализации. Требуется, однако, общее решение, и здесь мы имеем пример задачи, которую лингвистика ставит перед теорией графов.

Выше мы рассмотрели лишь несколько примеров простейшего применения теории графов в лингвистике. Речь идет не только о новом описании известных вещей; в теории графов находим возможность новой поста-

новки некоторых задач, связанных с формой языка; мы находим также понятия, методы и результаты, позволяющие решить некоторые из этих задач и ставящие некоторые новые задачи, которые мы не могли бы даже сформулировать без этой теории.

Для более полного знакомства с теорией графов отсылаем читателя к работам [2], [13] и [14].

Механико-математический факультет Бухарестский Университет

ЛИТЕРАТУРА

1. J. Apostel, B. Mandelbrot, A. Morf. Logique, langue et théorie de l'information, Presses Universitaires de France. Paris, 1957.
2. Claude Berge. Théorie des graphes et ses applications. Dunod, Paris, 1958.
3. J. Cantineau. Les oppositions significatives. «Cahiers Ferdinand de Saussure» 10, 1952, стр. 11—40.
4. J. Cantineau. Le classement logique des oppositions. «Word», vol. 11, N1, 1955, стр. 1—9.
5. N. Chomsky — G. Miller. Finite state languages. «Information and Control», 1, 1958, стр. 91—112.
6. Р. Л. Добрушин. Элементарная грамматическая категория. «Бюллетень Объединения по проблемам машинного перевода», № 5. М., 1957, стр. 19—21.
7. P. Erdös. Some remarks on the theory of graphs. «Bulletin of the American Mathematical Society», vol. 53, 1947, стр. 292—294.
8. Z. Harris. Methods in Structural Linguistics. «University of Chicago Press», Chicago, 1951.
9. L. Hjelmslev. Prolegomena to a theory of language. Baltimore, 1958. «The international journal of American linguistics», vol. 19, №1 Suppl., memoir 7.
10. О. С. Кулагина. Об одном способе определения грамматических понятий на базе теории множеств. «Проблемы кибернетики», 1, 1958, стр. 203—214.
11. S. Marcus — Em. Vasiliu. Mathématiques et phonologie. Théorie des graphes et consonantisme de la langue roumaine. «Revue de mathématiques pures et appliquées», t. 5, fasc. 2, 1960, стр. 319—340, fasc. 3—4, 1960, стр. 681—704.
12. Ю. С. Мартемьянов. [Устное сообщение, ноябрь, 1960).
13. Г. Н. Поваров. О матричном анализе связей в частично ориентированных графах. «Успехи математических наук», 1956, т. XI, вып.: 5 (71).
14. Г. Н. Поваров. О структурной теории сетей связи. «Проблемы передачи информации», вып. 1. М., Изд-во АН СССР, 1959.
15. И. И. Ревзин. Некоторые проблемы формализации синтаксиса. «Бюллетень Объединения по проблемам машинного перевода», № 21. М., 1957, стр. 5—35.
16. Th. Skolem. Ein kombinatorischer Satz mit Anwendung auf ein logisches Entscheidungsproblem. «Fundamenta Mathematicae», t. 20, 1933, стр. 245—261.
17. N. S. Troubetzkoy. Principes de Phonologie. Paris, 1957.
18. С. К. Шаумян. Лингвистические вопросы кибернетики и структурная лингвистика. «Вопросы философии», 1960, № 9.
19. С. К. Шаумян. Двухступенчатая теория фонемы и дифференциальных элементов. «Вопросы языкознания», 1960, № 5, стр. 18—34.

[GRAPH-THEORY, LINGUISTIC OPPOSITIONS AND INVARIANT STRUCTURE

Summary

The classification of linguistic oppositions on the basis of inclusion relation among sets (as carried out by Trubetzkoy and Cantineau) does not contain but a part of the types of oppositions that are studied in modern linguistics. Some types of oppositions that fail to appear in the Trubetzkoy — Cantineau classification may be described by means of the order relation or the equivalence relation. But even these relations do not cover all linguistic necessities. A mathematical theory which furnishes a unitary description of all linguistic oppositions is the graph theory. This theory not only permits an advantageous description, but may also be used to raise and solve some problems concerning the structure of language. Examples in this direction are given in the paper, some aspects of the isomorphism between phonological and grammatical structures being thus pointed out.

Ю. К. Лекомцев

К ВОПРОСУ ОБ АНАЛОГИЯХ В СТРОЕНИИ СХЕМ СЛОГА И ПРОСТОГО ПРЕДЛОЖЕНИЯ

Настоящая работа посвящена свойствам, общим для звуковой и смысловой сторон языка. Эта проблема возникла в современном языкознании. Действительно, сведение современным языкознанием языковой системы к сложной сети отношений, членящих субстанцию, означает переход от изучения двух неоднородных явлений (звуковых колебаний и сложного психического отношения — значения) к изучению однородного явления — зависимостей, связей. Проблема закономерностей, общих для звуковой и смысловой сторон языка, является частью более общей проблемы аналогичного строения различных областей языкового целого, так называемой проблемы языкового изоморфизма. Мы ставим перед собой задачу прояснения общей проблемы языкового изоморфизма на примере сравнения конкретных языковых объектов. Сами эти объекты будут описаны в специально выбранной для этого системе понятий.

Непосредственная задача нашей работы заключается в анализе двух частных объектов (слога и предложения определенных языков) и в констатации свойств, общих для этих объектов (в первую очередь), и их специфических свойств. Свойства объектов, вскрытые при анализе, могут носить как общеязыковой, так и частный характер. Безусловно общеязыковыми должны оказаться те свойства, которые могут быть выведены из предпосылок какой-либо формальной теории. Однако могут, по-видимому, существовать общеязыковые свойства, невыводимые из существующих теорий языка (примером такого общеязыкового свойства можно считать обязательное наличие в языке типа слогов с начальным согласным) [1].

Свойства, невыводимые из формальной теории, могут быть признаны универсальными (общеязыковыми) или частными только на основе исчерпывающей эмпирической проверки на материале различных языков.

Так как мы не проводили проверки рассматриваемых нами свойств на материале различных языков, то часть выводов (и констатируемых свойств), а именно, утверждения, наименее связанные с предпосылками структурной теории [2], следует считать действительными лишь в пределах взятого нами материала.

Слог и предложение могут рассматриваться с различных точек зрения, на разных уровнях абстракции и т. д.

Определим более конкретно объект нашего анализа.

---

[1] «There are languages lacking syllables with initial vowels and or syllables with final consonants but there are no languages devoid of syllables with initials consonants or of syllables with final vowels». R. J a k o b s o n. Typological studies and their contribution to historical comparative linguistics. Report for the VIII International congress of linguists. Oslo, 1957, стр. 7.

[2] Речь идет о принципах. Для нас несущественно, что в настоящий момент структурная теория языка еще слабо формализована.

Когда люди говорят, они не могут свободно, произвольно выбирать и размещать более простые образования — элементы, в пределах более сложных образований — единиц; они не могут как угодно переставлять фонемы в слоге, морфемы в слове, слова в предложении, простые предложения внутри сложного предложения и т. д., и они не могут помещать любой из элементов на определенное место в соответствующей единице.

В некоторых случаях такие перестановки и взаимозамены могут оказаться допустимыми и не изменят желаемого действия наших слов, в других случаях перестановки и взаимозамены по-прежнему будут давать допустимые в данном языке, но уже другие единицы; наконец, в определенных случаях, такие перестановки и взаимозамены элементов приведут к образованию бессмысленных с точки зрения носителей языка последовательностей. Таким образом, нас интересуют принципы размещения элементов, взятых из одного и того же ограниченного списка, в пределах образуемых ими единиц. Совокупность операций, позволяющих из заданного списка элементов построить список единиц, включающий лишь определенные комбинации элементов, мы назовем схемой соответствующего типа единиц (схемой слога, слова, предложения и т. д.). Среди такого рода операций большую роль играют операции, приводящие к ограничению теоретически возможного числа комбинаций элементов (сравни понятие ограничения разнообразия в кибернетике).

Схема сложной языковой единицы представляется нам в виде определенного порождающего устройства [3], абстрактной кибернетической машины, правда, машины с довольно специфичным видом преобразований, о чем будет сказано ниже. На входе такая машина имеет элементы (например, фонемы в схеме слога, слова в схеме предложения) и в результате ряда операций на выходе машины возникают единицы (например, слоги и предложения).

Так обстоит дело при синтезе единиц. При анализе единиц на элементы процесс будет обратный.

Такие схемы, представляющие собой системы отношений между элементами и операций над ними, взятые с точки зрения синтеза единиц из элементов, вместе с самими наборами элементов и являются объектами нашего исследования. Таким образом, наши объекты представляют собой первичные данные, записанные на некотором метаязыке (а если говорить точнее на мета-метаязыке, так так наша запись имеет дело с уже готовыми понятиями, такими как фонема, позиционный вариант фонемы, фонологический слог, синтагматическая и парадигматическая ось, звуковой ноль и т. д., — не имеет значения тот факт, что в ряде случаев мы передаем эти понятия другими терминами). В дальнейшем нам придется различать сходство, лежащее в самих объектах, и сходство, связанное с нашим мета-языком.

Материалом работы для схемы слога послужили слоги классического тибетского языка и для схемы предложения — предложения вьетнамского языка. Слог классического тибетского языка является удобным объектом для исследования вследствие богатства групп согласных и вследствие его совпадения со звуковой стороной либо морфемы, либо слова (это упрощает вопрос о выделении слога). В настоящей работе, чтобы не входить в рассмотрение сложных вопросов исторической фонетики, мы исходим из упрощенного, более того, явно неверного представления, что состав фонем классического тибетского языка совпадает с составом

[3] В связи с этим укажем на чрезвычайно перспективную, на наш взгляд, работу И. А. Мельчука «О термине «система» в лингвистике» (рукопись находится в печати), где система любых языковых элементов рассматривается как порождающий механизм и где предлагается ряд количественных показателей, характеризующих механизм такого рода. Схема в нашем понимании близка в некоторых отношениях к системе в понимании И. А. Мельчука.

графем. Исходным для нас является письменное изображение слога. Конечно, особенности письма, носящие чисто орфографический характер, в расчет не принимались (например, употребление ◌ — «а чунг» после гласного). Слоги классического тибетского языка были взяты по словарю H. A. Jäschke .Tibetan-English Dictionary. London, 1881.

При рассмотрении вьетнамского простого предложения также были сделаны некоторые упрощения: так, было сокращено число парадигматических классов путем сведения нескольких групп модальных частиц в один класс, были также сведены в один класс косвенные дополнения совместности, инструментальности и эргативности. Для статистического подсчета было взято 600 простых предложений вьетнамского языка из современных художественных текстов. Наша работа строится следующим образом: сначала дается анализ слога в тибетском языке, затем анализ простого предложения во вьетнамском (кратко) и, наконец, отмечаются черты сходства и различия между схемами обеих единиц этих языков.

ОПИСАНИЕ СХЕМЫ ТИБЕТСКОГО СЛОГА

На входе: Состояние схемы представлено множеством занумерованных (т. е. упорядоченных) скобок [4], связанных отношением конъюнкции [5]: $(\quad)_1 \,\&\, (\quad)_2 \,\&\, (\quad)_3 \ldots$

Схема имеет 9 (или 12) скобок такого рода (см. примечание 6, стр. 34).

Дано множество элементов (в данном случае фонем):
k, *k'*, *g*, ŋ, *ć*, *ć'*, *'j*, *ñ*, *t*, *t'*, *d*, *n*, *p*, *p'*, *b*, *m*, *ts*, *ts'*, *dz*, *w*, *ź*, *z*, ○, *y*, *r*, *l*, *ś*, *s*, *h*, *'*, *a*, *i*, *u*, *e*, *o*, *ai*, *au*, *ea*, *eu*, *iu*, 0 (нуль).

Кроме элементов, манифестирующихся в звук, множество фонем содержит пустое подмножество (состоящее из единственного элемента — нуля, который не манифестируется в звук).

---

[4] Понятие скобок мы не определяем; это понятие в своей основе связано с позициями фонем (с позиционными планами фонем) в реальных слогах.

[5] Для записи отношений между элементами и группами элементов мы используем логические функции. Логические функции являются функциями истинности. Пусть каждый из двух объектов A и B будет охарактеризован либо как истинный (обозначение 1), либо как ложный (обозначение 0). Тогда конъюнкцией (обозначение &) называется такое отношение между A и B, которое истинно в том случае, когда истинны и A и B:

A	B	A&B
1	1	1
1	0	0
0	1	0
0	0	0

Логической суммой (или просто суммой) называется такое отношение между A и B, которое истинно только тогда, когда истинен только один из элементов:

A	B	A ‖ B
1	1	0
1	0	1
0	1	1
0	0	0

Дизъюнкцией (V) называется такое отношение между A и B, которое истинно тогда, когда истинно по крайней мере либо A, либо B:

A	B	A V B
1	1	1
1	0	1
0	1	1
0	0	0

Логические понятия истинности или ложности понимаются в нашей работе как «существование» или «несуществование» в языке данных единиц.

Этап I

Осуществляется операция размещения элементов по скобкам.

В зависимости от способа размещения все элементы множества распадаются на три подмножества (непересекающихся):

1) пустое подмножество — 0,

2) гласные фонемы — m_1,

3) согласные фонемы — m_2.

0 размещается во все скобки, кроме скобки 10; m_1 помещается только в скобку 10; m_2 распределяется по всем скобкам, кроме скобки 10, при соблюдении следующего правила. Элемент e_n из m_2, входящий в скобку c_n, не может входить в скобки c_{n-1} и c_{n+1}, т. е. любой элемент из m_2 может входить в несколько скобок. Элементы, находящиеся внутри скобок, мы рассматриваем как неупорядоченные.

На первом этапе наша схема будет иметь следующий вид (с учетом примечания [6]):

(B, 0)$_1$ & (R, *s*, 0)$_2$ & (*g*, *m*, *o*, *l*, 0)$_3$* & (*d*, 0)$_4$ & (*k*, *k'*, *g*, *p*, *ć*, *ć'*, *'j*, *ń*, *t*, *t'*, *n*, *p*, *p'*, *b*, *m*, *ts*, *ts*, *dz*, *ź*, *z*, *ś*, *s*, 0)$_5$ & (*h*, 0$_6$* & (*w*, *y*, *r*, *l*, 0)$_7$ & (*h*, 0)$_8$ & (*ɔ*, 0)$_9$* & (*a*, *i*, *u*, *e*, *o*, *ai*, *au*, *ea*, *eu*, *iu*)$_{10}$ & (*g*, *p*, *d*, *n*, *b*, *m*, *r*, *l*, 0)$_{11}$ & (*s*, 0)$_{12}$.

Совокупность элементов, размещенную в пределах одной скобки, назовем классом (К).

На первом этапе все скобки схемы представлены собственно классами: K_1 & K_2 & K_3* & K_4 & K_5 & K_6* K_7 & K_8 & K_9* & K_{10} & K_{11} & K_{12} (1).

Этап 2

Осуществляется операция разделения классов на подклассы, сопровождающаяся введением новой функции (суммы).

Каждый класс (K_n) разделяется на непустое подмножество (E_n) и нуль (0_n), связанные функцией суммы. E_n и 0_n мы назовем подклассами. K_{10} (класс гласных), не содержащий нуля, записывается как E_{10}.

Параллельно с введением суммы внутрь скобок происходит введение суммы между отдельными группами конъюнктивных скобок.

---

[6] При проверке схемы на данном этапе материалом оказывается, что не все согласные фонемы могут быть размещены по скобкам. Для некоторых согласных фонем нельзя указать точно позицию по отношению к другим фонемам. Так фонемы *r*, *s* встречаются перед фонемами *k*, *k'*, *g* и др., а фонемы *b* встречается как перед *k*, *k'*, *g* и др., так и перед *r*, *s*. Однако фонема *g*, *m*, *o* встречаются перед фонемами того позиционного класса, к которому принадлежат *k*, *k'*, *g* и др., но не сочетаются ни с *b*, ни с *r*, *s*.

Таким образом, позиция фонем *g*, *m*, *o* оказывается неопределенной относительно *b*, с одной стороны, и *r*, *s*, с другой.

Рассматривая явления такого рода, можно либо считать, что для подобных элементов (как *g*, *m*, *o*) утеряна часть информации об их связи со схемами кода, либо рассматривать эти явления как закономерные черты кодовых схем.

Условным приемом записи элементов такого рода может быть фиктивное отнесение подобных элементов к определенным скобкам, или же введение их в схему со времени состояния входа в виде скобок, связанных с некоторыми группами конъюнктивных скобок отношением «суммы».

Ради простоты записи, с одной стороны, и ради выделения элементов типа *g*, *m*, *o*, с другой, мы предпочитаем ввести в схему дополнительные скобки, соединенные с остальными скобками знаком конъюнкции, давая им особую отметку в отличие от скобок обычного типа (скажем, звездочку(*) над правой скобкой).

Скобки «со звездочкой» будут введены в схему в следующем порядке: элементы, для которых нельзя указать вхождение в скобки с номерами от 1 до N, будут вводиться в специальную скобку с номером N+1.

Сам факт несочетаемости определенных групп фонем будет отражен на следующем этапе.

Формула 1 приобретает следующий вид:

$\{[E_1 \| 0_1) \& (E_2 \| 0_2) \& 0_3^* \& (E_4 \| 0_4) \& 0_5 \& 0_6^* \& (E_7 \| 0_7) \& 0_8 \& 0_9^*] [E_1 \vee E_2 \vee E_4 \vee E_7]^7 \& [E_{10} \& [E_{11} \| 0_{11}) \& (E_{12} \| 0_{12}]\} \|$

$\{(E_1 \| 0_1) \& (E_2 \| 0_2) \& 0_3^* \& 0_4 \& E_5 \& 0_6^* \& (E_7 \| 0_7) \& 0_8 \& 0_9^* \& E_{10} \& (E_{11} \| 0_{11}) \& (E_{12} \| 0_{12})\} \|$

$\{0_1 \& 0_2 \& 0_3^* \& 0_4 \& 0_5 \& 0_6^* \& E_7 \& E_8 \& 0_9^* \& E_{10} \& (E_{11} \| 0_{11}) \& (E_{12} \| 0_{12})\} \|$

$\{0_1 \& 0_2 \& E_3^* \& (E_4 \| 0_4) \& 0_5 \& 0_6^* \& (E_7 \| 0_7) \& 0_8 \& 0_9^* \& E_{10} \& (E_{11} \| 0_{11}) \& (E_{12} \| 0_{12})\} \|$

$\{0_1 \& 0_2 \& E_3^* \& 0_4 \& E_5 \| 0_6^* \& (E_7 \| 0_7) \& 0_8 \& 0_9^* \& E_{10} \& (E_{11} \| 0_{11}) \& (E_{12} \| 0_{12})\} \|$

$\{0_1 \& 0_2 \& 0_3^* \& 0_4 \& 0_5 \& E_6^* \& (E_7 \| 0_7) \& 0_8 \& 0_9^* \& E_{10} \& (E_{11} \| 0_{11}) \& (E_{12} \| 0_{12})\} \|$

$\{0_1 \& 0_2 \& 0_3^* \& 0_4 \& 0_5 \& 0_6^* \& 0_7 \& 0_8 \& E_9^* \& E_{10} \& (E_{11} \| 0_{11}) \& (E_{12} \| 0_{12})\} = 1 \quad (2)$

Начиная с этапа 2, благодаря введению функции суммы, схема слога начинает порождать последовательности. На втором этапе ею (схемой) порождаются типы слогов [9].

Для того, чтобы схема начала порождать реальные слоги (в том виде как они представлены в словаре), необходимо перейти к разбиению непустых подклассов на более дробные группы элементов и фиксировать линейные комбинаторные возможности этих групп. Соответственно 0_n также оказывается разделенным на целый ряд нулей. Например, на этом этапе схема слога должна будет отразить отсутствие слогов *rar*, *rir*, *rur*, *ror* (но *rer* существует!), и существование слогов *ral*, *ril*, *rul*, *rol* (но *rel* отсутствует!) и т. д.

В настоящей работе, однако, мы не даем формулы (3): построить формулу (3) чрезвычайно трудоемкая задача; эта формула и пояснения к ней заняли бы слишком много места, хотя принципы ее построения не отличаются существенно от принципов построения формулы (2). Кроме того, мы все равно не смогли бы дать аналогичной формулы (3) для предложения, ввиду огромного числа групп (субподклассов), на которые должны быть разбиты непустые подклассы E_n. Следует заметить, что хотя формула (2) является заведомо избыточной, если ее рассматривать как механизм порождения конкретных слогов, все же она отражает наиболее существенное в структуре слога: формула (3), которая должна отразить все ограничения, накладываемые на сочетаемость фонем в слоге, несет с собой много случайного и кроме того отражает части структуры, наименее устойчивые в историческом развитии языка.

Однако мы дадим формулировку этапа 3.

Этап 3

Осуществляется операция разделений подклассов E_n, 0_n на субподклассы e_{n_1}, e_{n_2}, e_{n_3}..., o_{n_1}, o_{n_2}, o_{n_3}... Субподклассы распределяются по группам конъюнктивных скобок, связанных функцией суммы. Возможно также введение других функций в качестве дополнительных средств записи.

---

[7] Данная группа дизъюнкций отражает тот факт, что наличие начального согласного в тибетском слоге обязательно.

[8] Функция суммы между группами конъюнктивных скобок говорит о несочетаемости некоторых непустых подклассов E_n.

[9] Типом слога можно назвать результат замены каждого элемента реального слога языка соответствующим непустым подклассом E_n. Например тип слога E_1 E_2 E_5 E_7 E_{10} E_{11} (ср. реальный слог bsgrag) может быть получен из формулы (2), после того как мы выберем 2-ю группу конъюнктивных скобок, отбросив 1, 3, 4, 5, 6, 7 группы, соединенные функцией суммы, и представим в ней K_{12} нулем, а K_1, K_2, K_7 и K_{11} подклассами E_1, E_2 E_7 и E_{11} (см. формулу (2)).

ОПИСАНИЕ СХЕМЫ ВЬЕТНАМСКОГО ПРОСТОГО ПРЕДЛОЖЕНИЯ

Покажем, что понимается под элементом схемы вьетнамского простого предложения. Под «словом» во вьетнамском языке мы понимаем группу морфем (минимальных двусторонних знаков), включающую ряд служебных морфем, с одной стороны, и основу, состоящую из одной или нескольких неслужебных морфем, с другой (под служебными морфемами понимаются морфемы, образующие парадигмы с малым числом членов и не допускающие перестановки относительно основы; неслужебные морфемы образуют парадигмы с относительно большим числом членов).

Таким образом, под «словами» мы понимаем единицы такого рода, которые Ш. Балли назвал «синтаксическими молекулами» [10].

Некоторые разряды «слов» могут, однако, состоять только из основ (простых или сложных). Отсутствие служебных морфем служит одним из отличительных признаков «слов» этого рода.

Примером «слов» первого рода могут служить «слова-существительные» и «слова-глаголы» во вьетнамском языке. Например: *Trong*^{1} *nhữ'ng*^{2} *cái*^{3} *con*^{4} *ngu'ò'i*^{5} (*này*) 'Среди (этих) пятерых людей', где морфема с индексом 1 представляет предлог (или аналитический падеж), с индексом 2 — число (множественное), с индексом 3 — артикль, с индексом 4 — классификатор, а морфема с индексом 5 представляет основу; *cũng*^{1} *sẽ*^{2} *không*^{3} *muốn*^{4} *trốn*^{5} *đi*^{6} 'также в будущем не захотеть убежать', где морфема с индексом 1 выражает совместность действия, с индексом 2 — время (будущее), с индексом 3 — отрицание, морфема с индексом 4 есть основа модального глагола *хотеть*, с индексом 5 — основа глагола *бежать*, морфема с индексом 6 выражает направление действия (удаление).

Примером «слов» второго рода могут служить наречия (onomatopœia), например *khúc khích* 'исподтишка' (смеяться), модальные частицы: например: *ạ* — морфема, значением которой является уважение к слушающему.

«Слова» первого рода, включающие различные служебные морфемы, но идентичные основы, можно назвать «словоформами». «Слова» второго рода имеют только по одной «словоформе». «Слова-существительные» могут иметь при себе группы морфем, выступающие в роли определения. «Слова» первого рода (реже также и «наречия») могут образовывать цепи так называемых «однородных членов». Практически необходимо рассматривать однородные элементы такого вида как один элемент одной и той же парадигмы, одной и той же скобки.

Таким образом, мы выделяем в качестве особой единицы «словоформу», которая в некоторых случаях может быть распространена определением. Эту единицу можно назвать условно «звеном». Любые однородные «звенья» рассматриваются в качестве одного «звена».

Предложение и его элементы интересуют нас в их смысловом аспекте. Под элементом схемы простого предложения понимается сумма смыслов всех морфем, входящих в «звено». Предложение также рассматривается как сложная единица, состоящая из смыслов отдельных «звеньев».

При описании схемы вьетнамского простого предложения условно порядок «звеньев» в предложении и элементов в схеме рассматривался как абсолютно жесткий.

На входе. Состояние схемы представлено множеством занумерованных (т. е. упорядоченных) скобок, связанных отношением конъюнкции: $(\quad)_1 \,\&\, (\quad)_2 \,\&\, (\quad)_3 \ldots$

---

[10] См.: Ш. Балли. Общая лингвистика и вопросы французского языка. М., 1955, стр. 317.

Схема имеет 9 (или 11) скобок[11] такого рода.

Дано множество элементов (см. выше), включающее также пустое подмножество (ноль):

$$a_1\ a_2,\ a_3, \ldots\ a_n,\ 0.$$

Этап 1

Осуществляется операция размещения элементов по скобкам.

В зависимости от способа размещения все элементы множества распадаются на 3 подмножества: 1) пустое подмножество — 0, 2) подмножество m_1, 3) подмножество m_2.

0 размещается во все скобки. Элементы из m_1 размещаются только по определенным скобкам (1, 4, 6, 7, 8, 10, 11), причем каждый элемент входит только в одну скобку. Элементы из m_2 размещаются по определенным скобкам (2, 3, 5, 9), причем каждый элемент может входить либо в группу скобок 2, 5, 9, либо в группу скобок 3, 5, 9 (в одну из скобок или в 2, 3 скобки одновременно).

Множество элементов, входящих в скобку с индексом n, называется классом (K_n).

Схема вьетнамского простого предложения может быть представлена на первом этапе следующей формулой:

$K_1\ \&\ K_2\ \&\ K_3\ \&\ K_4\ \&\ K_5\ \&\ K_6\ \&\ K_7\ \&\ K_8\ \&\ K_9\ \&\ K_{10}^*\ \&\ K_{11}^* = 1\ (1)$[12].

Этап 2

Осуществляется операция разделения классов на подклассы, сопровождающаяся введением новой функции (сумма).

Каждый класс (K_n) разделяется на непустое подмножество (E_n) и нуль (0_n), связанные функцией суммы. E_n и 0_n мы назовем подклассами.

Параллельно с введением суммы внутрь скобок происходит введение суммы между отдельными группами конъюнктивных скобок.

Формула (1) приобретает вид:

$$(E_1 \| 0_1)\ \&\ (E_2 \| 0_2)\ \&\ (E_3 \| 0_3)\ \&\ (E_4 \| 0_4)\ \&\ (E_5 \| 0_5)\ \&\ (E_6 \| 0_6)\ \&\ (E_7 \| 0_7)\ \&\ (E_8 \| 0_8)\ \&\ (E_9 \| 0_9)\ \&\ 0_{10}^*\ \&\ 0_{11}^*) \| (0_1\ \&\ 0_2\ \&\ 0_3\ \&\ 0_4\ \&\ 0_5\ \&\ 0_6\ \&\ 0_7\ \&\ 0_8\ \&\ 0_9\ \&\ E_{10}^*\ \&\ 0_{11}^*) \| (0_1\ \&\ 0_2\ \&\ 0_3\ \&\ 0_4\ \&\ 0_6\ \&\ 0_7\ \&\ 0_8\ \&\ 0_{10}^*\ \&\ E_{11}^* = 1$$ [13] (2)

На данном этапе схема простого предложения порождает последовательности, соответствующие (при учете ряда упрощений) реальным

---

[11] Как и в случае со слогом, есть реальные элементы языка, для которых нельзя точно указать вхождение в одну из 9 скобок. Это два типа междометных предложений (на нашем материале). Для одного из них не определено вхождение ни в одну из 9 скобок, для другого — не определено вхождение в группу скобок с 5—9.

Для обоих типов мы вводим дополнительные скобки, отмеченные звездочкой *, соответственно 11 и 10.

[12] K_1 символизирует класс так называемых «модальных частиц» (он мог бы быть помещен в схеме также и на предпоследнем месте), K_2 — «обстоятельства времени» (беспредложная именная группа), K_3 — «обстоятельства места» (беспредложная именная группа), K_4 — «обстоятельства места» (именная группа с предлогом), K_5 — «подлежащее» (беспредложная именная группа), K_6 — «косвенные дополнения совместности, инструментальности и эргативности», K_7 — «наречия», K_8 — «сказуемое», K_9 — «прямое дополнение» (именная группа без предлога), K_{10}^* — «звукоподражательное междометие», K_{11}^* — «эмотивное междометие».

[13] Если при записи формулы (2) исходить непосредственно из 600 разобранных предложений, то формула (2) примет более сложный вид. Например, необходимо будет записать факты такого рода: невозможно присутствие E_5 и E_6 без E_9 и т. д. Но поскольку такие добавочные ограничения связаны лишь с данным конкретным материалом, мы не ввели их в формулу.

простым предложениям вьетнамского языка, с точки зрения их смысловой стороны, в которых каждый элемент заменен соответствующим непустым подклассом (E_n).

Следующий этап должен привести к порождению реальных простых предложений вьетнамского языка с точки зрения их смысловой стороны, представленных в качестве последовательностей элементов.

Этап 3

Осуществляется операция разделения подклассов E_n, O_n на субподклассы e_{n_1}, e_{n_2}, e_{n_3}, . . ., o_{n_1}, o_{n_2}, o_{n_3}. . . Субподклассы распределяются по группам конъюнктивных скобок, связанных функцией суммы. Возможно также введение других функций в качестве дополнительных средств записи.

Как и в случае со слогом, мы не даем формулы (3).

Формула (3) должна воспроизводить тончайшие семантические связи.

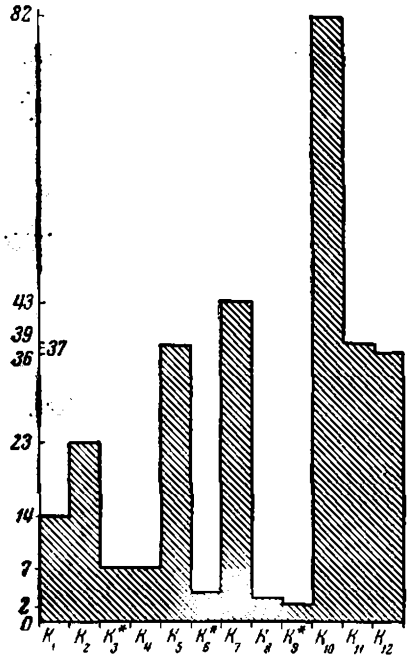


Рис. 1. Представление классов подклассами в схеме слога

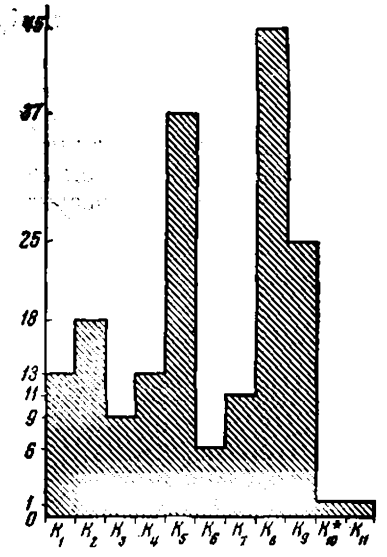


Рис. 2. Представление классов подклассами в схеме предложения

Перейдем к выявлению общих моментов в обеих схемах и затем к выявлению различий между ними.

Общими чертами схем слога и простого предложения являются следующие.

1) Обе схемы представляют собой порождающие устройства с элементами на входе и с последовательностями на выходе.

2) Числа этапов предположительно одинаковы (предположительно потому, что третий этап в обоих случаях не был проведен).

3) В обеих схемах в основе лежат одни и те же логические функции (конъюнкция и сумма).

4) В обеих схемах на каждом этапе производятся идентичные операции.

Интересно указать на присущий обеим схемам специфический характер операций, который отличает данные порождающие устройства от большинства других кибернетических машин: в наших схемах осуществляются не операции, преобразующие один элемент в другой элемент, а операции, организующие индивидуальные элементы в «класс как целое», и операции, делящие классы последовательно на подклассы различных ступеней.

5) В обеих схемах существуют скобки (классы), подклассы, субподклассы со звездочкой, введенные для таких элементов, вхождение которых в какую-либо конкретную ординарную (без звездочки) скобку (класс, подкласс, субподкласс) не определено однозначно.

6) В обеих схемах есть классы, которые на выходе наиболее редко бывают представлены нулем (K_{10} — класс гласных — для слога и K_8 — глагольное сказуемое — для предложения).

В связи с тем, что «центральным» элементам (гласный, основа, глагольное сказуемое и др.) уделяется большое внимание при рассмотрении различных аспектов «языкового изоморфизма», мы подкрепим последнее положение подсчетами.

Рассмотрим выходные последовательности второго этапа: здесь каждый класс — K_n представлен одним из своих подклассов — либо E_n либо 0_n.

Схема слога дает 82 последовательности из 7, 6, 5, 4, 3 и 2 членов [14].

Схема предложения дает 55 последовательностей из 6, 5, 4, 3, 2 и 1 членов.

В схеме слога класс K_n представлен нулем в N последовательностях из 82:

K_n	N	K_n	N
K_1	— 14	K_7	— 43
K_2	— 23	K_8	— 3
K_3*	— 7	K_9*	— 2
K_4	— 7	K_{10}	— 82
K_5	— 37	K_{11}	— 39
K_6*	— 4	K_{12}	— 36

В схеме предложения класс K_n представлен нулем в N-последовательностях из 55:

K_n	N	K_n	N
K_1	— 13	K_6	— 6
K_2	— 18	K_7	— 11
K_3	— 9	K_8	— 45
K_4	— 13	K_9	— 25
K_5	— 37	K_{10}*	— 1
		K_{11}*	— 1

Для наглядности изобразим представления классов K_n через подклассы E_n и 0_n в графиках.

Незаштрихованная часть указывает на число случаев с 0_n, а заштрихованная часть — на число случаев с E_n.

Подсчеты такого рода оправдывают сопоставление гласных и глагольного сказуемого, конечных согласных и дополнения (классы, наиболее часто представленные E_n среди классов с индексами, большими, чем индекс «центрального» класса); что касается сопоставления с подлежащим, то в нашем примере с ним не может быть сопоставлен класс начальных

[14] Практически мы брали конкретные слоги (в пределах словаря) и конкретные предложения (в пределах материала) и заменяли элементы последовательностей соответствующими E_n.

согласных K_5, но с ним может быть сопоставлен класс K_7 (*y*, *w*, *r*, *l*) — класс, наиболее часто представленный E_n среди классов с индексами меньшими, чем индекс «центрального» класса.

Следует заметить, что на графике схемы тибетского слога картина осложнена тем, что числовые показатели для E_n в некоторых классах (K_1, K_2, K_{12}) относительно высоки в силу грамматических функций фонем, представляющих эти классы (план выражения грамматических морфем иногда состоит из одной фонемы).

Важно разделить эти черты сходства на черты, сходные в силу их принадлежности к «аксиоматике и синтаксису» метаязыка, и черты сходства самих объектов.

Положение 1 было задано «аксиоматикой» нашей метатеории; очевидно, что оно также указывает на сходство объектов, но сходство этого рода не подлежит доказательству.

Положения 2, 3, 4, 5 и 6 не были заданы самим метаязыком. Действительно, если бы мы строили нашу теорию строго формально, мы, очевидно, в качестве отношений нашего исчисления взяли бы все типы функций, которые можно определить по таблице истинности. Тогда мы увидели бы, что выбор конъюнкции и суммы из общего числа функций в обоих случаях обусловлен свойствами объекта.

Точно также метатеория не содержит ограничений относительно числа этапов, совпадения типов отношений с этапами, наличия или отсутствия центрального класса и т. д.

Теперь следует обратиться к различиям между схемами. Эти различия велики. Создается впечатление, что основой качественных различий схем слога и предложения в значительной мере служат количественные различия — особенно различие в числе порождаемых схемами единиц и, соответственно, в числе исходных элементов — фонем и смыслов «звеньев».

Назовем основные, на наш взгляд, различия:

1) Число классов в схемах может быть различным (в нашем случае 12 и 11).

2) Число элементов на входе и последовательностей на выходе резко различно в обеих схемах.

3) Подмножества элементов 0, m_1 и m_2 распределяются по скобкам по-разному в обоих случаях.

4) В схеме предложения построение классов, аналогичных классам в схеме слога, делается возможным лишь при условии, что «однородные члены» рассматриваются как один элемент. Класс изображается нами как $a_n \| b_n \| B_n \| 0_n$, где a_n, b_n, B_n, 0_n субподклассы и $B_n = a_n \,\&\, b_n$.

Без наложения данного ограничения класс элементов предложения мог бы быть изображен как $(a_n \vee b_n) \| 0_n$.

5) Схема предложения обладает способностью усложнять свою структуру, вводя в качестве элементов некоторых своих классов целое предложение. Следует лишь вспомнить об определительных и дополнительных придаточных, о предложениях подлежащих и т. д., в то время как ни один класс в схеме слога не может быть представлен слогом.

6) Важным отличием схемы предложения от схемы слога является иная роль порядка элементов. Жесткий порядок подмножеств элементов, установленный нами для схемы простого предложения, является лишь грубым упрощением. Однако вопрос о роли порядка элементов в схемах чрезвычайно сложен и, возможно, имеет решающее значение для всей проблемы (проблемы моделирования языковых механизмов); мы не хотим поднимать этот вопрос в данной статье.

Из сравнения двух схем мы делаем следующие выводы: безусловно существует целый ряд общих закономерностей в построении обеих схем.

Существуют также значительные различия между схемами слога и предложения. Предполагается, что эти различия связаны в основном с числом исходных элементов (фонем, смыслов «звеньев»).

Что касается вопроса об «изоморфизме планов», то мы представляем его себе следующим образом. Между отдельными явлениями в пределах различных единиц языка, рассматриваемых на различных уровнях абстракции и т. д., действительно можно установить взаимно однозначные соответствия на основе тех или иных отношений. Например, по относительной высоте заштрихованных столбцов на графиках 1 и 2 можно сопоставить K_{10} — класс гласных, K_7 — класс плавных согласных корня, K_{11} — класс конечных согласных корня на первом графике с K_8 — классом глагольного сказуемого, K_5 — классом подлежащего, K_9 — классом прямого дополнения на втором графике. Взаимно однозначное отображение элементов двух множеств, сохраняющее отношение (или операции) между элементами, и называется, как известно, изоморфным отображением. Установление изоморфных отображений может служить в качестве приема исследования, причем приема искусственного и требующего дополнительных обоснований, но оно недостаточно для адекватного описания языковой структуры. При описании языковой структуры требуется, как нам кажется, учет двух факторов:

1) Наиболее общим случаем в языке является гомоморфное отображение (т. е. практически приходится сопоставлять различное число элементов).

2) Гомоморфизм (или изоморфизм как частный случай гомоморфизма) часто устанавливается между элементами языка на основе нескольких отношений одновременно. Так, например, удобно рассматривать элементы двух схем: A, B и A^1, B^1 как изоморфные на основе двух отношений — суммы и конъюнкции:

$(A \parallel B)_n \;\&\; (\bar{A} \parallel \bar{B})_{n+1}$ изоморфно $(A^1 \parallel B^1) \;\&\; (\bar{A}^1 \parallel \bar{B})_{n+1}$, где $\bar{A}$ означает элемент, отличный от A, а $\bar{B}$ — элемент, отличный от B. В данном примере присутствует и третье отношение — отношение порядка между индексами скобок. Итак, наиболее подходящим понятием для описания языковых связей было бы, как нам кажется, понятие гомоморфного отображения, взятого на основе нескольких отношений. Частным случаем такого гомоморфного отображения было бы изоморфное отображение, взятое по нескольким отношениям.

ЛИТЕРАТУРА

1. В. В. Иванов. Лингвистические вопросы создания машинного языка для информационной машины. Материалы по машинному переводу. Сборник I, изд. ЛГУ. Ленинград, 1958 (см. стр. 35—39).
2. В. Н. Топоров. Введение вероятности в языкознание и его последствия. Тезисы. Совещание по математической лингвистике 15—21 апреля 1959 г. Ленинград, 1959 (см. стр. 13).
3. С. К. Шаумян. Структурная лингвистика как имманентная теория языка. М., 1958.
4. С. К. Шаумян. Логический анализ понятия структуры языка. «Тезисы конференции по машинному переводу (15—21 мая 1958)». Москва, 1958, стр. 65—66.
5. J. Cantineau. Oppositions significative. «Cahiers F. de Saussure», t. 10, 1951, стр. 11—41.
6. C. Ebeling. Phonemics and functional semantics. «Lingua», vol. III, № 3, стр. 309—322.
7. L. Hjelmslev. Prolegomena to a theory of language. Baltimore, 1953.
8. J. Kuryłowicz. La notion de l'isomorphisme. TCLC, vol. V, Recherches structurales, 1949, стр. 48—60.
9. Linguistique et théorie de signe. «Journal de Psychologie», t. 41, N 2, 1949.
10. R. S. Pittman. Nuclear structures in linguistics. «Language», vol. 24, No 3, 1948, стр. 287—293.
11. L. J. Prieto. Rapport paradigmatique et rapport syntagmatique sur le plan de contenu. «Omagiu lui Jorgu Jordan». București, 1958.

Summary

This paper presents an analysis of the problem of similarity in structure between two different units of language: the syllable and the sentence. The author has chosen as his material syllables of Old Tibetan and sentences of Vietnamese. The syllable and the sentence are considered as sequences genereted by certain devices (generators). A possible kind of generation is suggested. The two processes of generation (syllable and sentence generation) are compared. All the differences and similarities are summed up. In conclusion the author suggests that it is possible to establish homomorphism between language units of different character and level. Homomorphism may be based on more than one relation.

Р. Г. Пиотровский

О ТЕОРЕТИКО-ИНФОРМАЦИОННЫХ ПАРАМЕТРАХ УСТНОЙ И ПИСЬМЕННОЙ ФОРМ ЯЗЫКА

Тот факт, что во многих языках мира для передачи одной и той же семантической информации используется по крайней мере два алфавита — буквенный и фонемный (соотв. алфавит иероглифов и алфавит устных слов), —давно уже ставит перед языкознанием вопрос о соотношении этих алфавитов. Изучение этой проблемы имеет первостепенное практическое значение уже хотя бы потому, что почти каждое лингвистическое исследование требует установления соотношений между устной и письменной формами языка. Большинство языковедов, постоянно сталкиваясь с этой задачей, решает ее обычно, исходя из молчаливого соглашения о том, что между устной и письменной формами языка существует полный изоморфизм, опирающийся на систему прямых корреспонденций между символами устного и письменного алфавитов.

Между тем, общая система этих корреспонденций по существу не определена ни для одного языка мира. Существующие эмпирические правила преобразования устных символов в письменные регулируют лишь отношения между фонемами и буквами.

Еще сложнее обстоит дело со знаками препинания и просодическими элементами (словесное и фразово-синтагматическое ударение, интонации), которые также выступают в качестве элементарных символов буквенного и фонемного алфавитов. Здесь установлена лишь частичная эквивалентность между символами, ср. [?]≅[вопросительная интонация+ +пауза]; [!]≅[восклицательная интонация+пауза], [.]≅[утвердительная интонация+пауза], [;]≅[утвердительная интонация+короткая пауза] и, возможно, некоторые другие. Такие символы письменного (буквенного) алфавита, как большая буква, черточка, отчасти запятая, двоеточие и некоторые другие не имеют эквивалентов в устном (фонемном) алфавите, равно как словесное ударение и эмфатические долготы не имеют соответствий на письме. Интервалы в устной речи, разделяющие обычно синтагмы, не эквивалентны пробелам между словами письменной речи.

Цель настоящей статьи состоит в том, чтобы определить круг вопросов, которые должны быть в первую очередь рассмотрены при изучении общей проблемы изоморфизма и асимметрии, существующих между устной и письменной формами языка. Там, где это позволяет имеющийся в нашем распоряжении материал, мы предлагаем предварительное решение некоторых из поставленных вопросов.

Если исходить из положения о том, что язык, в какой бы форме он ни реализовался, представляет собой код, на который накладываются определенные вероятностные ограничения, а речь выступает в роли сообщения, реализующего этот код, то проблему соотношения устной и письменной форм языка следует рассматривать не только в плане сугубо

языковедческой методики, но также в сфере лингвистического приложения основных идей теории информации.

В связи с этим указанную проблему целесообразно рассматривать с точки зрения количественной оценки информации, передаваемой устной и письменной речью, а также с точки зрения количественного распределения информации в устном и письменном слове.

Говоря об измерении количества информации и ее распределении, следует различать понятия селективной и семантической информации. Применительно к языку селективная информация измеряется количеством неопределенности (энтропии) в системе выборов тех или иных языковых элементов (букв, фонем, слогов, морфем, слов, словосочетаний, предложений и т. д.) в данном участке речевой цепи. Семантическая информация измеряется количеством неопределенности, содержащейся в системе выборов событий, которые описываются языковым сообщением.

Применительно к языку селективная информация может использоваться для описания и измерения внутриязыковых связей. Семантическая информация является в этом случае экстралингвистической категорией. Как ни различны понятия селективной информации и семантической информации, между ними в области языка существует определенная связь и взаморегулировка [1] (ср., например, статистические зависимости между длиной слова, его частотой и полисемией). В настоящей статье соотношение устной и письменной речи рассматривается в плане селективной информации (в дальнейшем везде для краткости — просто «информация»).

* * *

Рассматривая соотношение устной и письменной речи с количественно-информационной точки зрения, мы будем использовать следующие теоретико-информационные понятия.

К о л и ч е с т в о и н ф о р м а ц и и (I), с о д е р ж а щ е е с я в с о о б щ е н и и, зависит от числа элементов сообщения (q) и основания кода (m), а также от распределения вероятностей этих элементов (т. е. от их полных вероятностей — p_i) и от наличия вероятностных корреляций между указанными элементами (т. е. от их условных вероятностей $p_i(j)$) [2].

Э н т р о п и я (H) определяется как мера неопределенности данного опыта. При этом количество информации, извлеченное из опыта a', будет характеризоваться величиной:

$$I(a') = H_{(a')}$$

В нашем случае под опытом a' мы будем понимать выбор символа (фонемы или буквы) a'.

О п т и м а л ь н а я э н т р о п и я (H_0) определяется как мера неопределенности выбора одного символа при m возможных равновероятных исходах (то есть при условии, что данный алфавит букв или фонем состоит из m равновероятных символов) и выражается следующей величиной:

$$H_0 = k\left(-\frac{1}{m}\log\frac{1}{m} - \frac{1}{m}\log\frac{1}{m} - \ldots - \frac{1}{m}\log\frac{1}{m}\right) = -k\log\frac{1}{m} = k\log m \quad (1)$$

Коэффициент пропорциональности k определяется основанием логарифма и выбором системы единиц, в которых измеряется информация.

---

[1] Ср.: I. Bar-Hillel. An Examination of Information Theory. «Philosophy of Science», vol. XXII, 2, 1955, стр. 92—105.

[2] Ср.: А. А. Харкевич. Очерки общей теории связи. М., 1955, стр. 94.

Поскольку мы впредь будем пользоваться логарифмами с основанием 2 и двоичными единицами измерения информации (битами), постольку

$$k = \frac{1}{\log_2 2} = 1.$$

Поэтому формулу (1) можно переписать следующим образом:

$$H_0 = \log_2 m \qquad (1a)$$

В связи с этим коэффициент k в формулах определения различных видов энтропии будет опускаться.

Энтропия отдельно взятой буквы (фонемы) (H_1). Поскольку буквы (фонемы) $l_1, l_2, l_3, \ldots, l_m$ ($f_1, f_2, f_3, \ldots, f_m$) каждого реального алфавита обладают различной вероятностью употребления ($p_1, p_2, p_3, \ldots, p_m$), формула (1) принимает следующий вид:

$$H_1 = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - p_3 \log_2 p_3 \ldots p_m \log_2 p_m = -\sum_{i=1}^{i=m} p_i \log_2 p_i \qquad (2)$$

Условная энтропия буквы (фонемы) ($H_1 j$) или (H_{II}, H_{III}, ... H_n). Буквы (фонемы) выступают в сообщении не только как неравновероятные, но и как коррелированные величины. В этом случае мы имеем дело не с полной, но с условной вероятностью появления той или иной буквы (фонемы). Смысл этой условной вероятности состоит в том, что вероятность появления данной буквы (фонемы) зависит от вероятности появления 1, 2, 3, ..., n непосредственно предшествующих букв.

Общая формула условной энтропии символа выглядит так:

$$H_n = -\sum_{i,\,j} p\,[b_i\,(n-1),\ j] \log_2 p_{b_i(n-1)}(j) \qquad (3)$$

где $b_i(n-1)$ обозначает сочетание из $(n-1)$ букв (фонем), предшествующих j-той букве (фонеме).

$p[b_i(n-1)]$, есть вероятность появления этого сочетания, а $p_{b_i(n-1)}(j)$ является условной вероятностью появления j-той буквы (фонемы) вслед за сочетанием из $(n-1)$ букв.

В связи с этим условная энтропия буквы (фонемы), коррелированной с одной предшествующей буквой (фонемой), то есть при условии расчета вероятности двухбуквенного сочетания, равна величине:

$$H_{II} = -\sum_{i,\,j} (p(i,\ j) \log_2 p_i(j) \quad \text{и т. д.} \qquad (3a)$$

Условная энтропия руквы (фонемы), вероятность которой обусловлена вероятностью двух предшествующих букв (фонем), равна величине:

$$H_{III} = -\sum_{h,\,i\,m\,j} p(h,\ i,\ j) \log_2 p_{(h,\,i)}(j) \quad \text{и т. д.} \qquad (3b)$$

Поскольку введение неравновероятности и коррелятивных ограничений употребления букв (фонем) уменьшает степень неопределенности их выбора, поскольку справедливы следующие отношения внутри бесконечного ряда энтропий:

$$H_0 \geqslant H_1 \geqslant H_{II} \geqslant H \geqslant_{III} \ldots, \geqslant H_n \text{ при } \to \infty^3$$

---

[3] Полное обоснование этих формул и математический комментарий к ним см.: C. E. Shannon and W. Weaver. The Mathematical Theory of Communication. Urbana, 1949; V. Belevitch. Langage des machines et langage humain. Bruxelles. 1956, стр. 13. А. А. Харкевич. Указ. соч., стр. 94—95; ср. Л. Бриллюэн, Наука и теория информации. Пер. с англ. М., 1950, стр. 19—28, 41—48; А. М. Яглом и И. М. Яглом. Вероятность и информация, изд. 2-е. М., 1960, стр. 187—195.

Избыточность письменной и устной форм речи (R). Код способен передать оптимальное количество информации при условии, что все входящие в него элементарные символы независимы, равновероятны и обладают в связи с этим максимальной неопределенностью ($I_0=H_0$). При этом условии код лишен какой-либо избыточности.

Введение вероятностных (полные вероятности) и вероятностно-корреляционных (условные вероятности) ограничений приводит к сокращению неопределенности в выборе символа (H_i) и одновременно к уменьшению количества информации, передаваемой одной буквой (фонемой) (I_i) по сравнению с оптимальным количеством информации, которое могло бы быть передано одним символом при отсутствии указанных ограничений. Таким образом $(H_i=I_i)<(I_0=H_0)$.

Раз общее число элементарных символов в коде (языке), на который наложены вероятностные ограничения, остается тем же, что и в коде, содержащем равновероятные и взаимонезависимые элементарные символы, а количество информации, передаваемой первым кодом, уменьшается по сравнению с объемом информации, которую можно передать с помощью второго кода, следует говорить об избыточности кода, несущего вероятностные ограничения. Эта избыточность определяется по формуле:

$$R_i=\frac{H_0-H_i}{H_0} \qquad (4),$$

где H_0 — оптимальная энтропия символа (то есть буквы или фонемы); H_i — полная или условная энтропия символа (то есть буквы или фонемы), вероятность которого обусловлена вероятностью $(i-1)$ предшествующих символов (см. выше, стр. 45); R_i — избыточность языка при учете вероятностных ограничений, распространяющихся на речевой отрезок длиной в i символов (букв или фонем).

Если бы в языке не существовало вероятностно-корреляционных ограничений, т. е. употребление всех символов было бы равновероятным и несвязанным, то избыточность равнялась бы нулю:

$$R_0=\frac{H_0-H_0}{H_0}=0. \qquad (4а)$$

Учет полных вероятностей букв (фонем) дает величину избыточности первого порядка R_I

$$R_1=\frac{H_0-H_I}{H_0}\geqslant 0. \qquad (4б)$$

Соответственно учет условных вероятностей букв (фонем) в двучленных сочетаниях дает величину избыточности второго порядка R_{II}

$$R_{II}=\frac{H_0-H_{II}}{H_0}\geqslant R_I\geqslant 0 \qquad (4в)$$

Аналогичным образом получаются величины, характеризующие избыточности третьего, четвертого и т. д. порядков. Избыточность монотонно растет по мере того, как учитываются новые условно-вероятностные ограничения. На определенном уровне она приближается к реальной избыточности языка, которая может быть определена как

$$R_k=\frac{H_0-H_k}{H_0} \qquad (4г),$$

Ср. также: И. М. Яглом, Р. Л. Добрушин, А. М. Яглом. Теория информации и лингвистика. «Вопросы языкознания», 1960, № 1, стр. 102—103; Е. В. Падучева. Возможность изучения языка методами теории информации. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста», вып. 4. М., 1961, стр. 9—10.

где k — максимальное расстояние (в буквах или фонемах) между двумя буквами (фонемами), на которое распространяются вероятностные ограничения в тексте.

При этом, $H_k \simeq H_{k+1} \simeq H_{k+2} \simeq \ldots \simeq H_n$. Это значит, что сочетание из (k—1) букв (фонем), предшествующих k-той букве (фонеме), в такой же степени определяет вероятность выбора этой последней, как сочетание из k предшествующих символов определяет вероятность появления (k+1)-го символа и т. д. В связи с этим и $R_k \simeq R_{k+1} \simeq R_{k+2} \simeq \ldots \simeq R_n$. Поскольку величина энтропии символа (буквы или фонемы) зависит от основания кода (m), то есть от числа символов, составляющих данный алфавит, постольку сопоставление величин энтропии в различных языках, стилях и формах функционирования языка возможно лишь при условии, что алфавиты сравниваемых языковых систем имеют одинаковое количество символов.

При сопоставлении различных языков и их разновидностей, обладающих алфавитами с разным количеством символов, необходимо пользоваться такими теоретико-информационными величинами, которые как бы приведены к общему знаменателю и практически мало зависят от величины m. Этому условию отвечают величины, характеризующие избыточность, которые, как это видно из формул (4) и (4а)—(4г), не зависят от величины m, то есть от количества букв или фонем, образующих данный алфавит.

В большинстве теоретико-информационных исследований рассматривается лишь один вид избыточности — так называемая реальная избыточность языка (R_k — см. формулу (4г)) [4].

Поскольку в лингвистических работах, использующих теоретико-информационные методы, приходится иметь дело с вероятностно-корреляционными связями различной длины (в основном это полные вероятности и условные вероятности символа в двух- и трехбуквенных — соотв. фонемных — сочетаниях), постольку необходимо ввести такие понятия, как избыточность первого, второго, третьего и т. д. порядков, что и было сделано выше.

В табл. 1 приведены данные об энтропии и избыточности разных порядков для письменной и устной форм некоторых европейских языков (часть этих данных уже приводилась в литературе, часть рассчитана заново).

Однако далеко не все из этих сведений могут быть использованы для выявления теоретико-информационных особенностей устной и письменной форм языка. Дело в том, что расчет данных относительно вероятности букв (фонем) и их сочетаний производился, исходя из различных лингвистических и теоретико-информационных условий. Во-первых, при определении энтропии и избыточности письменной речи в число букв включается пробел между словами, который не учитывается при обработке устной речи (здесь обычно принимается во внимание пробел между синтагмами, который имеет гораздо меньшую частотность, чем пробел между словами). Поскольку пробел между словами является «буквой», обладающей сравнительно большой вероятностью (до 20% в языках, в которых средняя длина слова равняется приблизительно 5 буквам), избыточность письменного текста должна быть выше избыточности устного текста — ср. в этом смысле данные таблицы 1 (колонки R_I, R_{II}, . . ., R_n)[5].

---

[4] Ср.: C. E. Channon and W. Weaver. Указ. соч., стр. 3—28; C. E. Shannon. Prediction and Entropy of Printed English. «The Bell System Technical Journal (BSTJ), vol. XXX, 1, 1951; С. Голдман. Теория информации, пер. с англ., М., 1957, стр. 57; Л. Бриллюэн. Указ. соч., стр. 44 и 46; А. М. Яглом, Р. Л. Добрушин, И. М. Яглом. Указ. соч., стр. 103.

[5] Ср. также: А. М. Яглом, Р. Л. Добрушин, И. М. Яглом. Указ. соч., стр. 106—107.

При составлении табл. 1 использованы собственные данные (итальянский язык), а также данные, заимствованные из следующих работ: Д. С. Лебедев, В. А. Гар-

Однако, было бы преждевременно говорить о более высокой избыточности письменной речи по сравнению с устной речью, исходя лишь из учета или неучета пробела. Во-первых, нельзя забывать о том, что при расчете энтропии и избыточности обеих форм функционирования языка не учитываются обычно вероятности использования знаков препинания (в письменной речи), а также просодических элементов, как интонация и ударение (в устной речи). Между тем введение в буквенный и фонемный алфавиты знаков препинания и, соответственно, просодических элементов может заметно изменить получаемую величину энтропии, а возможно, и избыточности [6].

Во-вторых, в связи с тем, что при расчете письменного текста принимается во внимание пробел между словами, а при обработке устного текста этот пробел не учитывается, нарушается корреспонденция между вероятностями буквенных сочетаний и вероятностями соответствующих им комбинаций фонем. Дело в том, что вероятность фонемосочетаний охватывает вероятность появления этих комбинаций как внутри слова, так и на стыках слов, в то время как вероятность буквосочетаний равна их вероятности лишь внутри слова [7].

В какой степени влияют эти различия в методах расчета вероятностей буквосочетаний и фонемосочетаний на числовые величины энтропии и избыточности письменной и устной речи, остается пока неясным.

В-третьих, при определении энтропии фонемы принципиально важно, производится ли расчет частотности фонемы из транскрибированного (или «произнесенного») книжного текста, т. е. текста того же жанра, что и текст, из которого были извлечены данные о вероятности и энтропии букв и буквосочетаний, или же этот расчет произведен по записям устной речи. В первом случае величина энтропии и избыточности зависит от двух

маш. Статистический анализ трехбуквенных сочетаний русского текста. «Проблемы передачи информации», вып. 2. М., 1959, стр. 79—80 (русский язык); С. Гольдман. Теория информации, пер. с англ. М., 1957, стр. 43 (английский язык); Г. Г. Белоногов. Некоторые статистические характеристики русских печатных текстов. «Доклады на межвузовской конференции по структурной и математической лингвистике». Черновцы, 22—28 сентября 1960 г. (русский язык); G. Dewey. Relative frequencies of English speech sounds. Harvard University Press. Harvard, 1923 (английский язык, устная форма); C. E. Shannon. Prediction and Entropy of Printed English. «Bell System Technical Journal», vol. XXX, I, 1951. стр. 50—56 (английский язык, письменная форма); E. C. Cherry, M. Halle, R. Jacobson. Toward the Logical Description of Languages in their Phonemic Aspect. «Language», vol. XXIX, 1, 1953, стр. 41—45 (русский язык, устная форма); K. Küpfmüller. Die Entropie der deutschen Sprache. «Fernmeldetechnische Zeitschrift», Jg. VII, 6, 1954, стр. 367—370 (немецкий язык); G. A. Barnard. Statistical Calculation of Word Entropies for Four Wester Languages. «IRE. Transactions on Information Theory», vol. I, 1, 1955, стр. 49—53 (немецкий язык); Ed. Nicolau, C. Sala şi Al. Roceric. Observaţii asupra entropiei limbii romîne. «Studii şi cercetări lingvistice», an. X, 1, 1959, стр. 39—53 (румынский язык).

[6] Так, по данным Г. Г. Белоногова (ср. его доклад, прочитанный 23 сентября 1960 г. на Черновицкой конференции по прикладной лингвистике), энтропия, приходящаяся на слово в русской деловой речи без учета знаков препинания, равна 10.62 дв. ед., а при учете знаков препинания эта величина уменьшается до 10.10 дв. ед. Ср. в этом плане данные из таблицы об избыточности первого порядка для румынского публицистического печатного текста: без учета знаков препинания (для 27 букв) R_1=12,2%; с учетом знаков препинания (для 42 букв) R_1=19%. Информация (а следовательно и ее количественная мера — энтропия), содержащаяся в просодических элементах, также достаточно велика. Ср. R. M. Fano. Communication Theory Point of View in Speech Communication. «Journal of Acoustical Society of America», vol. XXII, 6, 1950, ср. 691; Е. В. Падучева. Указ. соч., стр. 12—13; Ср. Б. Мандельброт. Закон Берри и определение «ударения». Сб. «Теория передачи сообщений», пер. с англ. М., 1957, стр. 248—254.

[7] В русском языке вероятность появления фонемного сочетания *ст* внутри слова равна 31,17%, а на стыках слов — 5,17%; вероятность появления сочетания *тр* внутри слова равна 8,06%, а на стыках слов — 2,16% и т. д. См.: Л. Р. Зиндер. О лингвистической вероятности. «Вопросы статистики речи (Материалы совещания)». Л., 1958, стр. 60.

переменных: во-первых, от своеобразий количественно-информационных структур фонемного и буквенного кодов языка, а во-вторых, от различий в методике расчета вероятностей букво-и фонемосочетаний. Действием этих двух переменных объясняются различия в величинах энтропии и избыточности в устной и письменной формах итальянского языка (см. табл. 1).

Во втором случае величина энтропии определяется действием не только указанных двух переменных, но и еще одной переменной. Существо этой переменной заключается в особом выборе лексики, фразеологии и синтаксических конструкций, характеризующих устно-разговорный и письменно-книжный стили языка [8]. Действие всех трех переменных определяет различия в величинах энтропии и избыточности устной и письменной форм русского и английского языков.

Для того чтобы исследовать взаимодействие всех этих переменных, нужно сначала проследить действие каждой из них в отдельности. Для этого необходимо провести такое сопоставление устной и письменной речи, чтобы получаемые при этом данные зависели от действия одной переменной, а другие переменные оказались бы устраненными. Современная техника лингвистического и теоретико-информационного эксперимента дает возможность проследить действие третьей переменной, которая, как уже указывалось, отражает различия в вероятностно-корреляционной организации (т. е. различия в выборе лексики, фразеологии, синтаксических конструкций) устно-разговорного и книжного (беллетристического) стилей. Для этого расчет энтропии и избыточности символов в обоих стилях необходимо производить исходя из использования одного и того же алфавита. Две первые переменные, связанные с особенностями структур буквенного и фонемного алфавитов, а также с различиями их использования, благодаря этому автоматически устраняются. Примером могут служить приведенные в табл. 1 данные об энтропии и избыточности символов в поэтическом, фольклорном, публицистическом жанрах румынского языка, с одной стороны, и в румынском народно-разговорном стиле, с другой. Эти данные получены в результате обработки текстов, записанных только буквенным алфавитом, включающим 26 символов (без учета пробела). К сожалению, указанные данные относятся лишь к значениям H_1 и R_1 и поэтому не могут служить основанием для каких-либо заключений о сравнительной избыточности устного и книжного стилей.

Более полные сведения об энтропии и избыточности беллетристического и устно-разговорного стиля дают эксперименты по последовательному отгадыванию букв (фонем, слов и т. д.) в связном тексте [9].

В нашем распоряжении имеются сведения об энтропии беллетристического и устно-разговорного стилей в русском и румынском языках. Эти сведения получены путем проведения упрощенного эксперимента по отгадыванию букв в 132 стобуквенных текстах (по 33 текста на каждый стиль в каждом из указанных языков) [10].

Данные эксперимента были обработаны по формуле

$$H_x = H_1(1 - P_0 - P_1) + P_1$$

---

[8] Подробнее о лингвистических различиях между понятиями «устная форма языка» и «разговорный стиль речи» см.: Р. Г. Пиотровский. Очерки по стилистике французского языка, изд. 2-е, Л., 1960, стр. 18—20.

[9] Методика проведения и типы экспериментов подробно описаны в следующих работах: C. E. Shannon. Указ. соч., стр. 53 и сл.; Л. Бриллюэн. Указ. соч., стр. 48—50; А. М. Яглом, Р. Л. Добрушин, И. М. Яглом. Указ. соч., стр. 105—106; Е. В. Падучева. Указ. соч., стр. 16—20.

[10] Данные по русскому языку взяты из работы «Энтропия русского языка» (в выполнении которой участвовал автор настоящей статьи). Отчет о докладе на эту тему, прочитанном в Институте языкознания АН СССР, см. в журнале «Вопросы языкознания», 1961, № 4. Сведения по румынскому языку заимствованы из подготавливаемого совместно с Л. Л. Новак исследования, посвященного теоретико-информационным измерениям румынского языка.

Энтропия элементарного символа (буквы или

Язык	Форма (стиль)	Количество символов в алфавите (*основание кода m*)	H_0	H_1	H_{II}	H_{III}	H_V
Русский	Письменный	32 (с пробелом)	5	4,35	3,52	3,01	
		31 (без пробела)	4,95	4,46			
		47 (с пробелом и знаками препинания)	5,55				
	Устный	42	5,38	4,77	3,68	0,70	
Английский	Письменный	27 (с пробелом)	4,76	4,03	3,32	3,10	2,16
		25 (без пробела)	4,70	4,14	3,57		
	Устный	42	5,38	4,71			
Немецкий	Письменный	27 (с пробелом)	4,76	4,10			
		26 (без пробела)	4,70	4,097		2,80	
Итальянский	Письменный	22 (с пробелом)	4,46	4,00			
		21 (без пробела)	4,39	3,96			
	Устный	31	4,95	4,35			
Румынский	Письменный:						
	поэзия	26 (без пробела)	4,70	4,21			
	фольклор	26 » »	4,70	4,39			
	публицистика	26 » »	4,70	4,114			
	То же	27 (с пробелом)	4,76	4,119			
	»	42 (со знаками препинания)	5,38	4,345			
	Устный:						
	народно-разговорная речь	26 (без пробела)	4,70	4,16			

где x — номер буквы, H_x — энтропия буквы с заданным номером, H_1 — энтропия первого порядка в данном языке, вычисляемая по формуле (2), P_0 — вероятность того, что появление буквы x полностью предопределено цепочкой из x—1 предшествующих букв, P_1 — вероятность того, что буква x будет отгадана с первой попытки, причем имеются и другие возможности выбора буквы, разность $(1 - P_0 - P_1)$ представляет собой, таким образом, вероятность того, что буква x будет отгадана с двух и более попыток. В результате были получены четыре модели распределения количества информации (энтропии) в стобуквенных текстах беллетристического и устно-разговорного жанров русского (табл. 2) и румынского языков (табл. 3). Путем преобразования по формуле (4) величин, характеризующих энтропию в каждой из указанных моделей, получаем распределение избыточности в каждом из рассматриваемых стилей (рис. 1, 2).

Как видно из рисунков, устно-разговорный стиль в обоих языках менее избыточен, чем книжный (беллетристический) стиль. В русском языке избыточность устно-разговорного стиля приближается к 68% (точнее, она колеблется в пределах 55—80%), а избыточность беллетристического

Таблица 1

фонемы) и избыточность в различных языках

H_{VI}	H_{VII}	H_{VIII}	H_n	R_0	R_I	R_{II}	R_{III}	R_V	R_{VI}	R_{VII}	R_{VIII}	R_n
		1,36		0	0,13	0,30	0,40				0,73	
	1,56			0	0,10					0,65		
		1,27		0							0,77	
				0	0,11	0,33	0,87					
	2,14	1,87	1,40	0	0,16	0,30	0,35	0,54		0,55	0,61	0,71
		2,35		0	0,12	0,24					0,50	
				0	0,12							
				0	0,14							
2,00			1,60	0	0,13		0,42		0,58			0,66
				0	0,11							
				0	0,10							
				0	0,12							
				0	0,10							
				0	0,07							
				0	0,125							
			1,30									0,72
				0	0,122							
				0	0,19							
				0	0,115							

стиля близка к 70% (точнее, колеблется между 59 и 79%). Соответственно для румынского языка эти величины равны 59% (47—70%) и 61% (60—67%).

Поскольку для устно-разговорного стиля характерна большая вариантность в выборе форм, чем в строго нормализованном книжном стиле, можно ожидать, что и в других языках избыточность последнего будет выше избыточности устно-разговорного стиля. Соответственно в устно-разговорном стиле будет приходиться на символ большее количество селективной информации, чем в книжном (беллетристическом) стиле.

* * *

Теперь рассмотрим вопрос о количественном размещении информации в письменном и устном слове.

Описанный выше эксперимент по отгадыванию букв был проведен и относительно отдельно взятых русских и румынских слов. В результате были получены модели распределения количества информации (энтропии) в русских и румынских словах различной длины (от двух до двенадцати

букв, включая пробел). Оказалось, что во всех случаях начальные буквы несут наибольшее количество информации (энтролия) — ср. в этом плане модели пятибуквенных и десятибуквенных слов (рис. 3, 4).

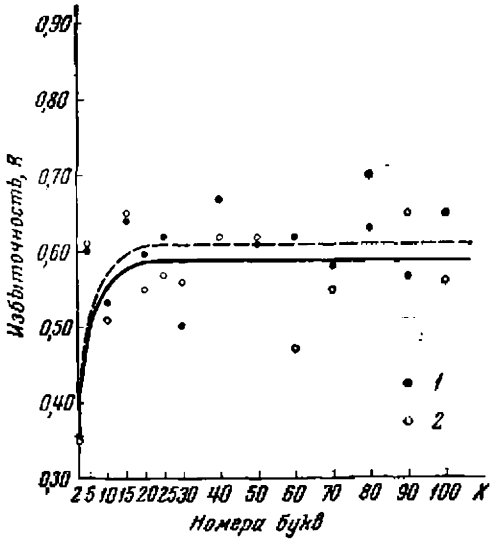


Рис. 1. График избыточности письменной формы румынского языка.
Стиль: *1* — разговорный; *2* — беллетристический

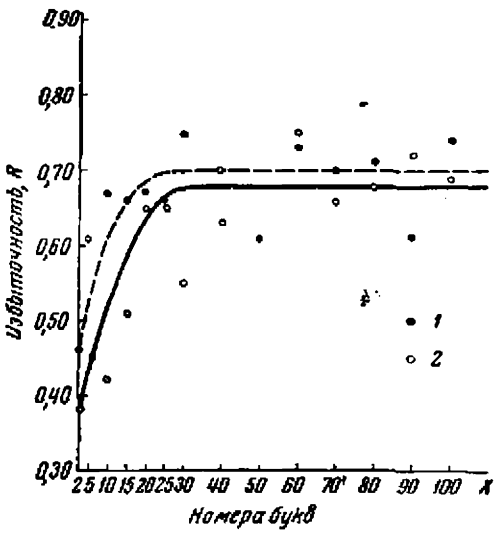


Рис. 2. График избыточности письменной формы русского языка
Стиль: *1* — разговорный; *2* — беллетристический

Аналогичных экспериментальных данных о распределении информации внутри письменного слова по другим языкам нет. Что же касается устной речи, то в настоящее время еще не разработана объективная методика, пользуясь которой можно было бы определить здесь распределение

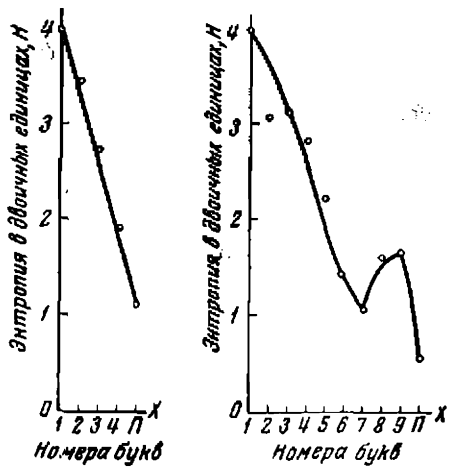


Рис. 3. Распределение энтропии в моделях пяти- и десятибуквенных румынских письменных слов

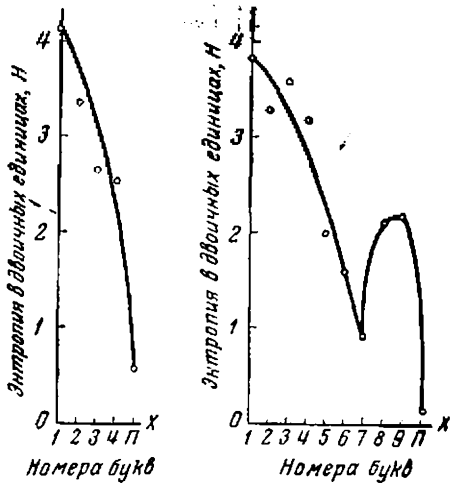


Рис. 4. Распределение энтропии в моделях пяти- и десятибуквенных русских письменных слов

информации. В связи с этим при дальнейшем рассмотрении интересующего нас вопроса мы вынуждены будем пользоваться методом прямого наблюдения и примерной оценки лингвистических фактов. Косвенным образом эти факты сигнализируют о внутренних, скрытых от прямого

Таблица 2

Выборочные данные о распределении энтропии и избыточности в моделях стобуквенных русских текстов разговорного и беллетристического стилей

Стили	X	P_0	P_1	H_x	R_x
Разговорный	1	—	—	4,22	0,16
	2	0	0,37	3,11	0,38
	5	0,18	0,49	1,93	0,61
	10	0,03	0,40	2,88	0,42
	15	0,15	0,37	2,46	0,51
	20	0,21	0,52	1,75	0,65
	25	0,33	0,34	1,77	0,65
	30	0,24	0,31	2,27	0,55
	40	0,18	0,52	1,83	0,63
	50	0,34	0,30	1,93	0,61
	60	0,36	0,46	1,24	0,75
	70	0,30	0,40	1,70	0,66
	80	*0,43*	*0,27*	*1,58*	*0,68*
	90	0,30	0,49	1,40	0,72
	100	0,27	0,49	1,53	0,69
Беллетристический	1	—	—	4,22	0,16
	2	0	0,49	2,71	0,46
	5	0,18	0,25	2,73	0,45
	10	0,27	0,46	1,63	0,67
	15	0,30	0,40	1,70	0,66
	20	0,24	0,49	1,66	0,67
	25	0,43	0,24	1,68	0,66
	30	0,37	0,45	1,23	0,75
	40	0,30	0,46	1,50	0,70
	50	0,24	0,40	1,97	0,61
	60	0,36	0,43	1,34	0,73
	70	0,40	0,33	1,50	0,70
	80	0,37	0,39	1,43	0,71
	90	0,34	0,27	1,97	0,61
	100	0,49	0,27	1,31	0,74

наблюдения количественно-информационных отношениях, существующих между отдельными частями устного и письменного слова.

Как известно, с конца XIX в. и особенно в XX в. в языках различных систем широко используется аббревиация слов. Охватывая в первую очередь административно-политическую и научно-техническую лексику, аббревиация осуществляется, как правило, в условиях письменной речи. В этой связи обращает на себя внимание тот факт, что подавляющее число современных письменных аббревиатур построено на сохранении начальных отрезков или заглавных букв слова[11]. Ср. интернац. *sing.*, русск. *сов-*

[11] В письменной речи используются также аббревиатуры, построенные на использовании начальной буквы в комбинации с буквами (как правило, согласными) из середины или даже конца слова. Ср. интерн. *Cd*(Cadmium), англ. *dn* (done, down), фр. *prp* (préposition), польск. *np* (naprzyklad), *ska* (spólka), русск. *д-р* (доктор). Однако выборочно-статистическое сопоставление обоих типов сокращений показывает, что аббревиатуры, использующие сочетания начальной буквы с буквой из середины или конца слова составляют не более 15—20% всех сокращений. О частном случае таких комбинированных аббревиатур см.: И. М. Б е р м а н. О вставочном типе словообразования. «Вопросы языкознания», 1959, № 2, стр. 104—107.

Таблица 3

Выборочные данные о распределении энтропии и избыточности в моделях стобуквенных румынских текстов разговорного и беллетристического стилей

Стили	X	P_0	P_1	H_x	R_x
Разговорный	2	0	0,33	3,07	- 0,35
	5	0,09	0,61	1,84	0,61
	10	0,12	0,42	2,31	0,51
	15	0,18	0,55	1,66	0,65
	20	0,21	0,36	2,12	0,55
	25	0,09	0,55	2,02	0,57
	30	0,12	0,49	2,09	0,56
	40	0.21	0,46	1,81	0,62
	50	0,03	0,70	1,81	0,62
	60	0,12	0,36	2,50	0,47
	70	0,15	0,43	2,15	0,55
	80	0,21	0,58	1,44	0,70
	90	0,09	0,67	1,65	0,65
	100	0,12	0,49	2,09	0,56
Беллетристический	2	0	0,33	3,07	1,69
	5	0,21	0,43	1,90	0,60
	10	0,09	0,49	2,21	0,53
	15	0,21	0,49	1,23	0,64
	20	0,24	0,40	1,87	0,60
	25	0,12	0,58	1,81	0,62
	30	0,15	0,36	2,36	0,50
	40	0.27	0,46	1,57	0,67
	50	0,18	0,49	1,84	0,61
	70	0,33	0,30	1,82	0,62
	70	0,24	0,36	2,00	0,58
	80	0,27	0,40	1,75	0,63
	90	0,18	0,43	2,03	0,57
	100	0,30	0,40	1,63	0,65

нарком, *ГЭС*; укр. *колгосп*; англ. *cl. ex.* (cloth extra), *Ho For* (Home Forces), фр. *C. G. T* (Confédération Générale du Travail); нем. *S.(üdl).* B.(*r*). (südlicher Breite), индонез. *FDR* (Front Demokrasi Rakjat); тур. *TDK* (Türk Dili Kurumy).

Концентрация информации в начале письменного слова и связанное с ней широкое использование начальной аббревиации в письменной речи не случайны. Написанное слово воспринимается как линейная последовательность составляющих его букв. Поэтому по мере продвижения от начала слова вероятность появления определенной буквы все возрастает. Соответственно уменьшается неопределенность в выборе буквы, то есть ее энтропия, которая, как это указывалось выше, выступает в качестве меры количества информации.

Отсюда становится ясным, почему наибольшую информационную нагрузку несут первые буквы написанного слова.

Иное положение в устной речи. Здесь наряду с отбрасыванием конца слова — процессом, характерным для письменных аббревиатур — ср. англ., фр. просторечно-фамильярное *doc* (doctor, docteur), *prof* (professor, professeur), англ. *cap* (captain), нем. *Schupo* (Schutzpolizei) — встре-

чается немало случаев аббревиации начальных слогов слова, в том числе — информационно наиболее нагруженной с точки зрения письменной речи — первой буквофонемы слова. Ср. англ. *bus* (omnibus), *cute* (acute), *ply* (apply), *spital* (Hospital), *tosh* (mackintosh, рядом с *mackin*), фр. арго *pitaine* (capitaine), *mandant* (commandant) [12], русск. разгов. *Ван-Ваныч* (Иван Иванович), *Сан-Саныч* (Александр Александрович).

Однако аббревиатуры в устной речи встречаются не настолько часто, чтобы, опираясь на них, можно было делать далеко идущие выводы. Поэтому следует обратиться к вопросу о фонетических деформациях, которые претерпевает слово в ходе исторического развития языка и в процессе межъязыкового контакта.

В этих случаях фонетическая переработка слова является результатом действия многих факторов. Одним из важнейших среди них является информационная нагруженность отдельных фонем.

Можно утверждать, что в большинстве языков мира наибольшей устойчивостью и информационной весомостью обладают фонемы ударенного слога. Что же касается начальных фонем устного слова — особенно если речь идет о неударенных слогах, — то они не имеют той отчетливой стабильности, которая характерна для наиболее «весомых» с информационной точки зрения начальных (особенно первой) букв письменного слова. Об этом свидетельствуют случаи потери начальных фонем, их метатеза, а также наращивание протетических звуков — явления, встречающиеся в различных языках мира, в том числе и в индоевропейских, для которых вообще характерна устойчивость начала слова[13]. Ср. русск. *лошадь, слон* < тюрк. *alaša at, aslan*,[14] русск. *он* ~ укр. *він*; фин. *tuoli* < нем. *Stuhl* или русск. *стул*; кабард. *haxše* < тюрк. *aqča*; бур.-монг. диал. *шен'ісе* < русск. *пшеница*, ср. чередования в монгольских говорах пров. Ганьсу (КНР) *фән* ~ *хон* ~ *он*, *фуда* ~ *хугума* ~ *гума*[15] или халх.-монг. *ачӣ*, *гла* < древн. монг. *havi*(*начи*), *ahula*[16]; инд. кечуа *allin*, *quepi* < исп. *gallina*, *espiga* (< лат. *spica*)[17]; корейск. *Курапха* < англ. (через поср. китайск.) *Europe*; осет. *ағд* ~ пехл. *haxt* ~ курд. *hêt*; осет. *akkag* < араб., перс. *ḥaqq*; осет. *alʃamblaj* ~ *ʃalæmbulaj*, нем. диал. (швабск.) *braxt* (< *gebracht*), *blibə* (< *geblieben*), *degt* (< *gedeckt*) и др.[18]

Среди индоевропейских языков ослабление и отпадение начальных фонем и слогов особенно часто наблюдается в албанском и румынском языках при устном освоении ими латинизмов (древний слой) — ср. лат. *amicus* > алб. *mic*; лат. *april* > ст.-рум. и диал. *prier*, алб. *prill*; лат. *excutere* > рум. *scoate*, алб. *shkund*; лат. *fossatum* > алб. *ʃshat*, ст.-рум. *ʃsat* > совр. рум. *sat* и др.[19] Одновременно указанные балканские языки широко используют протезу начальных согласных, что свидетельствует о фонетическом ослаблении (а косвенным образом и о потере информационности)

---

[12] Подробнее см.: М. М. Сегаль. О некоторых типах этимологических дублетов в современном английском языке. «Ученые записки ЛГПИИ им. А. И. Герцена», т. 157, ф-т англ. яз. Л., 1958, стр. 125—127. Ср.: H. Frei. La grammaire des fautes. Paris, 1929, стр. 124—125; H. Kjellman. Mots abrégés et tendances d'abréviation en français. Uppsala, 1920.

[13] Ср.: R. Gauthiot. La fin de mot en indo-européen. Paris, 1913, стр. 34—35; J. Vendryès. Recherches sur l'intensité initiale en latin. Paris, 1902.

[14] Подробнее об отпадении начального гласного при заимствовании в русский язык см.: В. И. Абаев. Из истории слов. «Вопросы языкознания», 1958, № 1, стр. 121.

[15] Другие примеры см.: Г. Д. Санжеев. Монгольские языки и диалекты. «Ученые записки Института востоковедения», IV. М., 1952, стр. 66—67.

[16] Там же, стр. 59—60.

[17] Подробнее см: Д. Э. Ибарра Грассо. Иероглифическая письменность индейцев Андского нагорья. «Вопросы языкознания», 1958, № 1, стр. 101.

[18] Сводку немецких примеров см.: В. М. Жирмунский. Немецкая диалектология. М.—Л., 1956, стр. 158.

[19] Подробнее см.: Р. Г. Пиотровский. Формирование артикля в романских языках. Выбор формы. Л., 1960, стр. 115—117.

последних, — ср. лат. *oleum*, *orbus*, *ovum*>алб. *vaj*, *verber*, *ve*; лат. *admissarius*, *arcus*>рум. *harmăsar*, *harc* (диал.), алб. *harmshuer* (*harmëshor*), *harc* и др.[20]

Говоря о своеобразии размещения информации в устной форме языка, следует также иметь в виду парадигматическую информационную иерархию фонем и типов фонем. В некоторых языках гласные и согласные имеют различную информационную нагрузку. Так, например, исследования Энкеля по разборчивости немецкой речи показали, что при устранении из слова согласных и сохранении вокального «костяка» слова его опознаваемость резко сокращается. Наоборот, сохранение консонантного скелета слова, предусматривающее устранение гласных, дает сравнительно высокий процент опознаваемости [21].

В известной степени это перекликается с данными немецкой диалектологии. В немецких говорах неударные гласные характеризуются большой неустойчивостью, что отражается в их взаимозаменяемости и редукции [22].

Опыты по пересадке русских звуков также обнаруживают более высокую опознаваемость согласных по сравнению с гласными [23]. Вместе с тем известно, что русские гласные (особенно неударные) обладают как в литературном языке, так и в говорах (особенно в акающих) большой неустойчивостью (ср. их вариантность вокруг сильной фонемы) и тенденцией к выпадению, в отличие от согласных (особенно не конечных)[24]. В связи с этим уместно вспомнить замечание А. М. Пешковского по поводу соотношения значения и звуковой структуры в корне *вод-а*: «Значение это может быть связано только с теми звуковыми элементами, которые есть во всех словах, т. е. не с *вóд-*, не с *вóт-* не с *вад-* и не с *вад'-*, не с *вэд-* и не с *вэд'-* в отдельности, а с неким экцерптом из этих элементов, с неким как бы алгебраическим «корнем», извлеченным из них, что схематически можно представить так: *в*+(*о* или *а* или Λ)+(*д* или *д'* или *т*)»[25]. Если рассмотреть все эти факты с точки зрения распределения информации, то можно предположить, что для русского языка свойственна информационная неравнозначность фонем (противопоставление согласных и гласных) и их позиций в слове (противопоставление начальных и ударенных позиций конечным). Немецкий материал также можно рассматривать с точки зрения количественно-информационного противопоставления согласных и гласных.

Информационная иерархия согласных и гласных обнаруживается и в других языках — она отражается в особенностях арабской графики, фиксирующей только согласные и долгие гласные, а может быть, и в сингармонизме финно-угорских и тюркских языков.

---

[20] Р. Г. Пиотровский. Указ. соч., стр. 118.

[21] F. Enkel. Über den Informationsgehalt der Geräuschlaute. «Nachrichtentechnische Zeitschrift» (Fernmeldetechnische Zeitschrift). Jg. IX, H. 11, 1956, стр. 493—498.

[22] В. М. Жирмунский. Немецкая диалектология, стр. 147—148 и 156—168.

[23] Ср.: М. Ф. Деркач. К вопросу о фонации, как отличительном признаке, лежащем в основе дифференцированного восприятия звонких и глухих согласных. «Научно-технический сборник ГСНИИ», 3, стр. 73—77. Н. И. Дукельский. Синтез русской речи из отдельных звуков и ее фонодальный состав. Там же, 6, стр. 6—8; Л. А. Чистович. Применение статистических методов к определению фонетической принадлежности индивидуального гласного звука. Сб. «Вопросы статистики речи». Л., 1958, стр. 33—34.

[24] Ср.: Р. И. Аванесов. Фонетика современного русского литературного языка. М., 1956, стр. 105—133; Его же. Очерки русской диалектологии. М., 1949, стр. 39 и сл., 122 и сл.

[25] Цит. по статье: А. Н. Гвоздев. Вопросы фонетики. Что дают три типа транскрипции. «Вопросы языкознания», 1958, № 6, стр. 85; там же см. аналогичные разбивки слов *дом*, *ход*, *док*, *молод*, *мясо*, *сарай*, *топор*, *доктор*, *очень*.

Распределение информации в устном и письменном слове зависит также и от того, рассматривается ли данное слово изолированно (т. е. в парадигматике языка) или в контексте (т. е. в синтагматике языка). Так, например, в ходе описанного выше эксперимента выяснилось, что в русском слове *хорошо*, взятом вне контекста, при последовательном отгадывании букв слева направо все шесть букв дают определенные системы выборов. Поэтому отгадка каждой буквы дает определенное количество информации, равное той энтропии, которую снимает каждый частный эксперимент (отгадка буквы). Таким образом, в письменном слове *хорошо*, взятом вне контекста, информация распределена (хотя, вероятно, и неравномерно) по всем шести буквам. В тех же случаях, когда слово *хорошо* находилось в контексте, оно отгадывалось во всех случаях по двум или трем первым буквам. Это свидетельствует о том, что при контекстном употреблении письменного слова *хорошо* информация целиком сконцентрирована в первой половине слова, в то время как вторая его половина оказывается избыточной.

Что же дает в этом отношении устная речь? Эксперимент с синтезированными звуками, поставленный Л. А. Варшавским, показывает, что полная иллюзия произнесения слова *хорошо*, взятого вне контекста (т. е. в парадигматике языка), возникает у аудиторов даже тогда, когда аппарат воспроизводит следующую последовательность звуков: [ха-шо] ([р] оказывается здесь избыточным).

Для того чтобы определить распределение информации в устном слове *хорошо*, употребленном в контексте, нами был проведен следующий эксперимент. Трем испытуемым была предложена магнитофонная запись контекста *мне хорошо видно*, причем в слове *хорошо* последовательно вырезывались отдельные участки. Оказалось, что все испытуемые немедленно опознавали и правильно повторяли весь контекст в тех случаях, когда были оставлены конечные звуки слова *хорошо* — [шо]. В тех случаях, когда на ленте оставлялась первая половина слова, [хар], опознание слова достигалось после определенной паузы, во время которой испытуемые обдумывали контекстные связи этого звукосочетания.

Таким образом, включение устного слова в контекст также вызывает перераспределение информации. В рассмотренном нами устном контекстном слове *хорошо* информационно нагруженным оказывается конец слова, опирающийся на ударный слог. Начало слова [хар] оказывается в этом случае избыточным.

Рассмотренный материал позволяет сделать следующие выводы:

1. Устно-разговорный стиль, обладающий большим количеством выборов лингвистических элементов, дает, очевидно, более высокий уровень энтропии и, таким образом, несет большое количество селективной информации, чем книжный (беллетристический) стиль. Что же касается соотношения уровней энтропии и избыточности устной и письменной форм языка, то этот вопрос может быть решен лишь после тщательного исследования информационной структуры буквенного и фонемного алфавитов в различных языках.

2. В устной форме языка информация размещена иначе, чем в письменной форме. Если в письменном слове наибольшую информационную нагрузку несут первые буквы, то в устном слове информация концентрируется вокруг фонем ударного слова. Это различие связано, очевидно, с разным характером восприятия лингвистических единиц устной и письменной форм языка. Письменное слово воспринимается как линейная последовательность элементарных символов (букв). Устное слово или по крайней мере его слоги воспринимаются целиком (симультанно).

3. В устной форме ряда языков обнаруживается информационная иерархия гласных и согласных, причем оказывается, что первые несут меньшую информационную нагрузку, чем вторые.

4. Количественное размещение информации в слове зависит и от того, рассматривается ли данное слово (или другая лингвистическая единица) изолированно или в контексте. Характер контекста также может влиять на размещение информации.

Таблица 4

Распределение энтропии в моделях пяти- и десятибуквенных русских слов

Длина слов	X	p_0	p_1	H_x
Пять букв	1	—	0,06	4,15
	2	—	0,28	3,41
	3	0,02	0,46	2,72
	4	0,02	0,50	2,59
	5 (пробел)	0,42	0,58	0,58
Десять букв	1	—	0,15	3,85
	2	—	0,30	3,35
	3	—	0,22	3,61
	4	—	0,34	3,21
	5	0,18	0,46	2,03
	6	0,24	0,64	1,16
	7	0,46	0,42	0,94
	8	0,18	0,43	2,13
	9	0,18	0,40	2,23
	10 (пробел)	0,85	0,15	0,15

Таблица 5

Распределение энтропии в моделях пяти- и десятибуквенных румынских слов

Длина слов	X	p_0	p_1	H_x
Пять букв	1	—	—	4,0
	2	0	0,20	3,48
	3	0	0,44	2,74
	4	0,06	0,62	1,93
	5 (пробел)	0,30	0,56	1,13
Десять букв	1	—	—	4,0
	2	0	0,33	3,08
	3	0	0,30	3,17
	4	0,06	0,33	2,83
	5	0,18	0,36	2,24
	6	0,12	0,70	1,44
	7	0,39	0,46	1,07
	8	0,33	0,37	1,60
	9	0,27	0,43	1,66
	10 (пробел)	0,55	0,42	0,54

ON THEORETICO-INFORMATIONAL PARAMETRES OF THE ORAL AND WRITTEN FORMS OF LANGUAGE

Summary

The purpose of the present study is to determine how much information is conveyed by the different parts of the written and oral word. Examining the experimental and historico-linguistic data we see that in writing word the information is condensed in the initial part of the word, in the oral word the information is concentrated on the intense syllable.

В. Ю. Розенцвейг и Л. М. Уман

К ПРОБЛЕМЕ ГРАММАТИЧЕСКОЙ ИНТЕРФЕРЕНЦИИ

В работах по структурной лингвистике, посвященных языковым контактам, термином «интерференция» обозначаются явления отклонения от языковых норм, которые появляются при регулярном использовании одним и тем же человеком (или коллективом людей) разных языков [1]. Как правильно отмечает У. Вейнрейх, «термин „интерференция" предполагает перестройку моделей (patterns), возникающую вследствие введения чужеродного элемента в структурно более организованные области языка, как то: в основу фонологической системы, в значительную часть морфологии и синтаксиса, а также в некоторые области словаря. . . Было бы чрезмерным упрощением говорить здесь о заимствовании или же лишь о добавлении к некоторой сумме»[2].

Явления интерференции рассматривались до сих пор главным образом на фонологическом уровне[3]. Менее изучена интерференция грамматическая.

Более того, сама возможность взаимопроникновения грамматических систем некоторыми учеными отрицалась[4]. Однако более глубокое изучение вопроса приводит к убеждению, что человек, пользующийся двумя языковыми системами, не может полностью изолировать не только фонемы и лексемы, но также и граммемы этих двух языков.

Грамматическая интерференция возникает при любом языковом контакте. В настоящей статье рассматривается интерференция, которую приходится преодолевать при обучении иностранному языку, в частности французскому. Такой интерференции, казалось бы, не должно быть, поскольку учащийся получает от учителя (учебника) сведения, необходимые для правильного пользования языком. Известно, однако, что это не так: несмотря на систематическое обучение, учащийся совершает грамматические ошибки. Их обычно пытаются предупредить так называемым сопоставительным рассмотрением двух языков, родного и иностранного. Но это сопоставление, как правило, ограничивается данной грамматической категорией, т. е. не учитывается структурная и дистрибутивная функция грамматической единицы в данных языках. Кроме того, сопоставительное рассмотрение некоторой языковой категории в двух разных языках не учитывает, что при контакте двух языковых систем

---

[1] См.: U. Weinreich. Languages in contact. New York, 1953, стр. 1; Einar Haugen. Bilingualism in the Americas. A Bibliography and Research Guide. Published by the American Dialect Society. November, 1956.

[2] Указ. соч., стр. 1.

[3] Классической здесь остается работа Е. Поливанова: «La perception des sons d'une langue étrangère». TCLP, N 4. Prague, 1931.

[4] См.: Lucien Tesnière. Phonologie et mélange de langues. TCLP, N 8. Prague, 1939.

возникает третья система, структура которой ими определяется. Подобная система строится в сознании человека, изучающего иностранный язык. Допускаемые им грамматические ошибки могут быть поняты и преодолены только если рассматривать их как явления, связанные с построением такой системы.

Настоящая статья ставит задачей описать процесс грамматической интерференции и определить основные ее типы.

ПРОЦЕСС ИНТЕРФЕРЕНЦИИ

В практике преподавания иностранного языка, особенно на начальном этапе обучения, приходится часто сталкиваться с типичными ошибками учащихся, которые мы обычно называем русизмами.

Рассмотрим, к примеру, типичные ошибки учащихся в области употребления французского поссесива: *Les parents aimaient* ***son*** *fils* (вместо *leur*). *Prenez* ***son*** *livre* (вместо *votre*). *Montre-moi* ***sa*** *dictée* (вместо *ta*). *Je n'ai pas apporté* ***ses*** *cahiers* (вместо *mes*). *Mon crayon n'est pas taillé, donne-moi* ***le sien*** (вм. *le tien*).

Ясно, что все эти ошибки совершаются под влиянием родного языка. Однако при более подробном анализе примеров возникает ряд вопросов:

1) Почему совершается ошибка, в то время как учащимся известно, что во французском языке выбор формы поссесива зависит от лица и числа обладателей?

2) Русскому языку также свойственно различение лица и числа обладателей, и для учащихся это не может представлять особой трудности. Почему же одна форма *свой* оказывается сильнее, чем формы *мой*, *твой*, *его—ее*, *наш*, *ваш*, *их*, вместе взятые?

3) Почему ошибочный выбор всегда падает на одну и ту же форму 3 лица единственного числа (обладателей)?

В русском языке система поссесива основана на противоположении форм личного притяжательного местоимения и возвратного притяжательного местоимения[5].

Личные притяжательные местоимения	Возвратные притяжательные местоимения
Мой, моя, мое, мои	
Твой, твоя, твое, твои	
Наш, наша, наше, наши	*Свой, своя, свое, свои*
Ваш, ваша, ваше, ваши	

Личное и возвратное притяжательные местоимения представляют собой члены одной оппозиции. Основанием для сравнения является присущий им общий признак — выражение отношений принадлежности. Признак, который их различает, противопоставляет, дифференцирует — наличие или отсутствие значения принадлежности предмета субъекту действия. Возвратное притяжательное местоимение выражает принадлежность предмета субъекту действия: *Я беру* ***свой*** *карандаш, вы берете* ***свой;*** личное притяжательное местоимение не выражает принадлежности предмета субъекту действия (предмет принадлежит другому лицу): *Я беру* ***ваш*** *карандаш, они берут* ***наш.***

Возвратное притяжательное местоимение, выражая всегда принадлежность предмета субъекту действия, не нуждается в особых формах, которые указывали бы на лицо и число обладателей, поскольку обладателем является сам субъект действия, выраженный уже в речи.

[5] О формах *его*, *ее*, *их* см. ниже, стр. 64.

Теперь обратимся к французскому языку. Подобной оппозиции, присущей русскому языку, во французском нет. Он не различает, принадлежит предмет субъекту действия или не принадлежит. В связи с этим не существует и особых форм личного и возвратного притяжательного местоимения. Однако во французском языке имеется своя оппозиция в системе поссесива: притяжательный детерминатив противопоставляется притяжательному местоимению.

Притяжательные детерминативы	Притяжательные местоимения	
Mon, ma, mes	*Le mien*	*les miens*
	La mienne	*les miennes*
Ton, ta, tes	*Le tien*	*les tiens*
	La tienne	*les tiennes*
Son, sa, ses	*Le sien*	*les siens*
	La sienne	*les siennes*
Notre, nos	*Le, la nôtre*	*les nôtres*
Votre, vos	*Le, la vôtre*	*les vôtres*
Leur, leurs	*Le, la leur*	*les leurs*

Основанием для сравнения является тот же общий признак — выражение принадлежности. Дифференциальный признак, по которому формы детерминатива и местоимения противопоставляются друг другу, основывается на том, выступает ли поссесив в качестве определителя существительного или заменяет существительное. Притяжательный детерминатив употребляется при наличии существительного, притяжательное местоимение — при отсутствии существительного: *Je prends **mon** crayon, vous prenez **le vôtre**.*

Итак, в русском и французском языках в одной и той же системе поссесива оппозиции различны. В русском языке вся совокупность значений поссесива разделяется на две группы: одна группа связана с формами личного притяжательного местоимения, другая — с формами возвратного притяжательного местоимения.

Во французском языке вся совокупность значений поссесива также делится на две группы, но имеющие иной характер: одна связана с формами притяжательного детерминатива, вторая — с формами притяжательного местоимения.

Русский язык

Личный	Возвратный
Поссесив	

Французский язык

Детерминатив	Местоимение
Поссесив	

Таковы основные оппозиции в системе поссесива указанных двух языков. Но, как известно, всякое грамматическое явление может обладать не одним дифференциальным признаком, а целым пучком признаков. Рассмотрим, какие признаки являются различительными для членов оппозиции, присущей русскому языку (табл. 1) и языку французскому (табл. 2). В этих таблицах плюс означает наличие, минус — отсутствие.

Таблицы показывают, что по целому ряду признаков внутри системы поссесива каждого языка имеются свои противопоставления.

Составим сводную таблицу по двум языкам, которая покажет, какие дифференциальные признаки могут быть внесены в третью систему (табл. 3).

Таблица 1

	Наличие признака		Наличие противопоставления
	личный поссесив	возвратный поссесив	
Значение принадлежности субъекту действия	—	+	+
Учет лица обладателей	+	—	+
Учет числа обладателей	+	—	+
Учет рода обладателей	—	—	—
Учет числа предметов	+	+	—
Учет рода предметов	+	+	—

Таблица 2

	Притяжательный детерминатив	Притяжательное местоимение	Противопоставление
Наличие существительного	+	—	+
Учет лица обладателей	+	+	—
Учет числа обладателей	+	+	—
Учет рода обладателей	—	—	—
Учет числа предметов	+	+	—
Учет рода предметов	+	+	—

Примеры ошибочного употребления поссесива во французском языке, приведенные выше, свидетельствуют о том, что учащиеся, используя систему выражения французского языка, вносят в систему содержания признаки, свойственные родному языку. Простейшая схема наложения языков позволяет наглядно представить область возможных ошибок в системе поссесива.

Французский язык

Детерминатив		Местоимение	
Личный	Возвратный *Область ошибок*	Личное	Возвратное *Область ошибок*

Внесение этой основной оппозиции из системы русского поссесива (личный — возвратный) влечет за собой внесение других дифференциальных признаков, по которым различаются члены данной оппозиции: выражение числа и лица обладателей.

Таблица 3

	Русский язык	Французский язык	Третья система
Противопоставление: возвратный поссесив — личный поссесив	+	—	+
Противопоставление: детерминатив — местоимение	—	+	+
Различение членов оппозиции по выражению: лица обладателей	+	—	+
числа обладателей	+	—	+
рода обладателей	—	—	—
числа предметов	—	—	—
рода предметов	—	—	—

Теперь мы можем ответить на вопросы, поставленные в начале статьи:

1) Почему совершается ошибка, в то время как учащимся известно, что во французском языке выбор формы поссесива зависит от лица и числа обладателей? Ошибка совершается из-за стабильности системы дифференциальных признаков родного языка в мышлении говорящего.

2) Почему форма «свой» оказывается сильнее, чем все личные формы, вместе взятые? Форма «свой», являясь членом оппозиции, противопоставлена всем личным формам, а в иностранный язык, как мы видели, оппозиция вносится целиком.

3) Почему ошибочный выбор всегда падает на форму 3 лица единственного числа (обладателей)? Иными словами, почему происходит ассоциация форм *свой*, *своя*, *свои* и *son*, *sa*, *ses*? Потому, что в русской системе поссесива имеется оппозиция: личный поссесив — возвратный поссесив. Возвратный поссесив имеет одну форму для всех лиц, тогда как личный поссесив имеет разные формы для разных лиц. Однако система личного поссесива не является полной: она обладает специальными формами только для 1 и 2 лица (*мой—наш*, *твой—ваш*). Специальных форм для 3 лица нет. В этой функции используются личные местоимения 3 лица в форме родительного падежа — *его*, *ее*, *их*, которые заимствуются системой поссесива из системы личного местоимения.

Надо полагать, что весь наш грамматический «репертуар» (термин Л. В. Щербы), т. е. все грамматические формы родного языка, существует в мышлении не в виде беспорядочного набора форм, а представляет собой упорядоченную систему различного рода оппозиций. Так, все личные формы существуют, очевидно, в виде трехчленных оппозиций, ибо в русском языке есть 1, 2 и 3 лицо, которые противопоставлены друг другу. Соответственно лицам в системе поссесива должны быть противопоставлены формы *мой—твой—его* (*ее*). По-видимому, это так и есть, потому что в родном языке эти формы употребляются правильно. Однако у большинства говорящих эта оппозиция недостаточно четко осознается из-за того, что формы 3 лица *его—ее* имеют свое законное место в системе оппозиций личного местоимения. В процессе языкового контакта, когда возникает потребность осознать местоположение формы *свой*

в системе поссесива, в мышлении возникает оппозиция *мой—твой—?*, лишенная третьего члена. Стремясь восстановить нарушенную оппозицию, мышление говорящего вносит в нее несвойственный ей член — форму *свой* (1—*мой*, 2—*твой*, 3—*свой*). (Часто на вопрос о том, какие притяжательные местоимения имеются в русском языке, учащиеся отвечают: *мой*, *твой*, *свой*.)

Таким образом форма *свой* наделяется несвойственными ей признаками — 3 лица единственного числа. По-видимому, эта ложная оппозиция и влияет на выбор эквивалента во французском языке — форм 3 лица единственного числа обладателей *son*, *sa*, *ses*.

Выбор данных форм поддерживается также и другими двумя факторами: а) обильным чтением книжных текстов, написанных в подавляющем большинстве в 3 лице, где формы поссесива 3 лица *son*, *sa*, *ses* встречаются значительно чаще, чем формы других лиц, а потому становятся наиболее привычными; б) возможно, в какой-то мере играет роль чисто внешнее созвучие форм *son*, *sa*, *ses* и *свой*, *своя*, *свои*.

Сказанное выше является лишним подтверждением того, насколько стабильны дифференциальные признаки родного языка в мышлении двуязычного человека. При внесении такого признака в иностранный язык, где не существует форм, которые различались бы по данному признаку, говорящий выделяет из общей системы какую-нибудь форму и противопоставляет ее остальным. Для выбора такой формы используется любая возможность, которая скрыта в родном языке, лишь бы сохранить внесенный в иностранный язык признак родного языка и как-то «узаконить» его существование в новой для него системе.

Итак, отклонение от нормы французского языка в области употребления поссесива происходит из-за внесения дифференциальных признаков родного языка в систему иностранного. Анализ других грамматических явлений также подтверждает мысль о том, что процесс грамматической интерференции состоит в замене системы грамматических дифференциальных признаков изучаемого языка другой, отклоняющейся от нее системой, построенной под воздействием системы грамматических дифференциальных признаков родного языка.

ТИПЫ ИНТЕРФЕРЕНЦИИ

Перенос дифференциального признака из родного языка в иностранный является лишь одним из возможных типов интерференции. В той же системе поссесива можно наблюдать на начальном этапе обучения еще один тип интерференции: в третью систему не вносится дифференциальный признак иностранного языка, оттого что он несвойствен родному языку.

Русский язык	Поссесив	
Французский язык	Детерминатив	Местоимение

Разделительная граница между детерминативом и местоимением уничтожается, и употребление детерминатива распространяется на область местоимения. Таким образом, третья система оказывается лишенной дифференциального признака, присущего системе иностранного языка.

	Русский язык	Французский язык	Третья система
Противопоставление: детерминатив — местоимение	−	+	−

Подобный тип интерференции часто наблюдается также в области употребления временных форм французского глагола. Рассмотрим, чем вызывается и как происходит процесс грамматической интерференции в системе глагольных форм прошедшего времени.

Представим себе всю совокупность значений прошедшего времени, выражаемых в русском языке глагольными формами прошедшего времени. В русском языке эти значения противопоставляются по в и д о в о м у признаку: а) значения завершенности и начинательности выражаются формой прошедшего времени совершенного вида; б) значения незавершенности, длительности, многократности — формой прошедшего времени несовершенного вида.

Русский язык

Прошедшее совершенное	Прошедшее несовершенное
Завершенность, начинательность	Незавершенность, многократность, длительность

Во французском языке эти значения противопоставляются прежде всего по в р е м е н н о м у признаку: а) значения прошедшего абсолютного; б) значения предшествования прошедшему. В первом случае употребляются passé composé и imparfait, во втором — plus-que-parfait (имеются в виду формы разговорной речи).

Французский язык

Passé composé, imparfait	Plus-que-parfait
Прошедшее	Предшествование прошедшему

Такая оппозиция несвойственна русскому языку. По временному признаку русские формы прошедшего времени не противопоставляются друг другу. Анализ ошибок учащихся свидетельствует о том, что данное противопоставление по временному признаку во французском языке учащимися не всегда соблюдается (проявляется тенденция к употреблению вместо plus-que-parfait форм passé composé, imparfait).

Русский язык	Формы прошедшего времени	
Французский язык	Passé composé, imparfait	Plus-que-parfait

Отсутствие данного дифференциального признака в русском языке ведет к уничтожению разделительной границы во французском языке и к распространению passé composé и imparfait на всю область значений прошедшего времени.

При помощи таблиц дифференциальных признаков рассмотрим образование третьей системы (табл. 4).

Сравнительная таблица показывает, что отсутствие временного признака (внутри системы прошедших времен) в русском языке тормозит перенесение его из французского языка в третью систему, ввиду чего последняя оказывается лишенной этого признака. С другой стороны, видовой признак русского языка переносится в третью систему, причем

Таблица 4

а) Русский язык

	Прошедшее	
	совершенное	несовершенное
Выражение завершенности	+	—
Противопоставление:		
трех лиц	—	—
двух чисел	+	+
по родам	+	+

б) Французский язык

		Plus-que-parfait
Выражение предшествования прошлому	—	+
Противопоставление:		
трех лиц	+	+
двух чисел	+	+
двух форм внутри одного члена оппозиции	+	—

в) Сравнительная таблица

	Русский язык	Французский язык	Третья система
Противопоставление:			
завершенности — незавершенности	+	—	+
прошлого предшествованию в прошлом	—	+	—
Различение членов оппозиции по выражению:			
лица	—	—	—
числа	—	—	—
рода	—	—	—
по количеству форм внутри членов оппозиции	—	+	—

этому способствует наличие двух форм — passe composé и imparfait — в первом члене французской оппозиции, которые различаются по видовому признаку (ниже об этом будет сказано более подробно).

Рассмотрим теперь каждый из членов оппозиции французского языка в отдельности.

Формы passé composé и imparfait образуют, в свою очередь, особую оппозицию, которая основана на видовом признаке. Но это не значит,

что можно поставить знак равенства между passé composé и прошедшим совершенного вида в русском языке или между imparfait и прошедшим несовершенного вида. Хотя в обоих языках дифференциальный признак один и тот же, распределение значений между членами оппозиции неодинаково. Русское прошедшее совершенного вида включает значения завершенности и начинательности, тогда как passé composé, помимо указанных значений, может выражать также ограниченную длительность (включая ограниченное многократное действие). Русское прошедшее несовершенного вида может иметь значения незавершенности, длительности, многократности, тогда как imparfait выражает длительность только неограниченную (включая повторяющееся действие, если количество повторений не указано).

Представим сказанное выше на схеме деления грамматических значений:

Русский язык

Прошедшее совершенное	Прошедшее несовершенное
Завершенность	Незавершенность

Французский язык

Passé composé	Imparfait
Завершенность + ограниченная длительность	Неограниченная длительность

или

Passé composé	Imparfait
Вся совокупность действия	Какой-нибудь период в процессе действия

Сравнение двух языков покажет точно область ошибок:

Русский язык	Прошедшее совершенное	Прошедшее несовершенное	
Французский язык	Passé composé	*Область ошибок (ограниченная длительность)*	Imparfait

При языковом контакте происходит как бы перемещение разделительной границы иностранного языка в направлении разделительной границы родного языка, в результате чего расширяется область значений у imparfait и соответственно сужается область значений у passé composé.

Русский язык			
Французский язык		←	

В данном случае интерференция проявляется в перераспределении объема значений у членов оппозиции. Как показывает практика, противодействовать данному типу интерференции особенно трудно.

Рассмотрим таблицу дифференциальных признаков (табл. 5).

Таблица 5

а) Русский язык

	Прошедшее	
	совершенное	несовершенное
Выражение завершенности	+	—
Противопоставление: трех лиц	—	—
двух чисел	+	+
по родам	+	+

б) Французский язык

	Passé composé	Imparfait
Выражение совокупности действия	+	—
Противопоставление трех лиц	+	+
Противопоставление двух чисел	+	+

в) Сравнительная таблица

	Русский язык	Французский язык	Третья система
Противопоставление: завершенности — незавершенности	+	—	+
Противопоставление: совокупности действия — процессу действия	—	+	—
Различение членов оппозиции по выражению: лица	—	—	—
числа	—	—	—
рода	—	—	—

Из сравнительной таблицы видно, что при наличии однородного дифференциального признака в обоих языках этот признак переносится в третью систему, но с распределением значений, свойственным родному языку. Это уже третий тип интерференции. Вернемся к нашей первой оппозиции:

Passé composé, imparfait — plus-que-parfait.

Внутри ее второго члена нет противопоставления, так как во французском разговорном языке нет других форм для выражения предшество-

вания прошлому (как известно, passé surcomposé не получило распространения в разговорной речи).

При контакте двух языков plus-que-parfait противопоставляется в русском языке оппозиции: прошедшее совершенное — прошедшее несовершенное:

Русский язык	Прошедшее совершенное	Прошедшее несовершенное
Французский язык	plus-que-parfait	

Поскольку в родном языке есть оппозиция, несвойственная иностранному языку, теоретически здесь создается возможность интерференции, т. е. возможность переноса дифференциального признака из родного языка в иностранный. Однако практически в данном случае интерференции не происходит (что подтверждается отсутствием ошибок у учащихся: если учащийся установил, что нужно выразить предшествование прошлому, он безошибочно употребляет plus-que-parfait, не путая его ни с какой другой формой).

Почему в данном случае интерференция нейтрализуется — этот вопрос будет рассмотрен ниже, при сравнении с futur simple.

В системе будущего времени может наблюдаться интерференция первых двух (рассмотренных выше) типов.

Представим совокупность значений будущего времени, выражаемых в русском языке формами будущего времени. Эти значения группируются (а следовательно и противопоставляются) в русском языке по в и д о в о м у признаку: завершенность выражается формой будущего времени совершенного вида, незавершенность — формой будущего времени несовершенного вида:

Русский язык

Будущее совершенное	Будущее несовершенное
Завершенность	Незавершенность

Во французском языке оппозиция основана на в р е м е н н о м признаке: будущее с точки зрения настоящего момента выражается формой futur simple, будущее с точки зрения прошлого — futur dans le passé.

Французский язык

Futur simple	Futur dans le passé
В плане настоящего	В плане прошедшего

Анализ ошибок свидетельствует о том, что данное противопоставление по временному признаку не всегда соблюдается учащимися. При контакте двух языков происходит интерференция, которая выражается в уничтожении дифференциального признака иностранного языка, в ликвидации разделительной границы и расширении области значений futur simple за счет futur dans le passé:

Русский язык	Формы будущего времени	
Французский язык	Futur simple	Futur dans le passé
		Область ошибок

Отсутствие дифференциального признака (временного) в русском языке ведет к уничтожению его во французском языке. С другой стороны, наличие в русском языке другого дифференциального признака (видового) может приводить к другого рода ошибкам, с которыми иногда приходится сталкиваться на начальном этапе обучения (буквальный перевод: *Я буду читать — Je serai lire*). Рассмотрим, что здесь происходит.

Французской форме futur simple противопоставлена в русском языке оппозиция: будущее совершенное — будущее несовершенное:

Русский язык	Будущее совершенное	Будущее несовершенное
Французский язык	Futur simple	

Оппозиция в русском языке основывается на видовом признаке (по содержанию). Кроме того, она основана и на различиях в формообразовании (что связано с системой выражения): будущее совершенного вида представляет собой форму синтетического образования (*прочту*), будущее несовершенного вида — форму аналитическую (*буду читать*). Это различие в формообразовании вызывает перенос дифференциального признака из русского языка во французский, что приводит к различению форм будущего времени во французском языке по видовому признаку и к образованию несуществующей там формы.

Рассмотрим табл. 6.

Таблицы показывают, что, во-первых, отсутствие в русском языке временного дифференциального признака мешает переносу его из французского языка в третью систему и, во-вторых, наличие в русском языке дифференциального признака по характеру формообразования влияет на перенос признака из русского языка в третью систему, что приводит к образованию несуществующей во французском языке формы.

Теперь вернемся к plus-que-parfait и сравним его с futur simple:

Русский язык	Прошедшее совершенное	Прошедшее несовершенное
Французский язык	Plus-que-parfait	
Русский язык	Будущее совершенное	Будущее несовершенное
Французский язык	Futur simple	

На первый взгляд примеры кажутся аналогичными. Однако в первом случае переноса видового признака не наблюдается, тогда как во втором случае данный признак иногда переносится. Чем это объяснить?

В русском языке как в одном, так и в другом случае противопоставление основано на видовом признаке. Имеется в виду дифференциальный признак, присущий системе содержания русского языка. С этой точки зрения примеры, приведенные выше, ничем не различаются. Сравним их с точки зрения системы выражения. Формы прошедшего совершенного и прошедшего несовершенного в русском языке различаются наличием или отсутствием префиксов и суффиксов (*написал—писал, записал—записывал*). В обоих случаях формообразование является синтетическим, и французский язык на него не реагирует. Нет у данных форм и других дифференциальных признаков, которые могли бы как-то поддержать

Таблица 6

а) Русский язык

	Будущее	
	совершенное	несовершенное
Значение завершенности	+	—
Противопоставление: трех лиц	+	+
двух чисел	+	+
Синтетическая форма	+	—

б) Французский язык

	Futur simple	Futur dans le passé
Выражение будущего в плане настоящего	+	—
Противопоставление: трех лиц	+	+
двух чисел	+	+
Синтетическая форма	+	+

в) Сравнительная таблица

	Русский язык	Французский язык	Третья система
Противопоставление: завершенности — незавершенности	+	—	+
Противопоставление: будущего в настоящем — будущему в прошедшем	—	+	—
Различение членов оппозиции по выражению: лица	—	—	—
числа	—	—	—
по характеру формообразования	+	—	+

различение членов оппозиции по видовому признаку. При контакте интерференция не происходит.

Другое дело во втором случае. Формы будущего совершенного и будущего несовершенного в русском языке различаются аналитическим и синтетическим характером образования. Системе выражения французского языка свойственны подобные противопоставления. Данный дифференциальный признак, присущий системе выражения русского языка, поддерживая видовой признак, свойственный системе содержания русского языка, приводит к переносу этого видового признака в систему содержания французского языка, что влечет за собой образование в системе выражения несуществующей во французском языке формы.

ВЫВОДЫ

1. При контакте двух языков, в частности при изучении иностранного языка, в мышлении говорящего происходит процесс замены системы грамматических дифференциальных признаков изучаемого языка другой, отличающейся от нее системой, построенной под воздействием системы грамматических дифференциальных признаков родного языка. Этот процесс называется грамматической интерференцией.

2. Анализ приведенных языковых фактов показывает, что грамматическая интерференция может быть следующих типов:

а) При наличии грамматического дифференциального признака в иностранном языке и отсутствии данного признака в родном языке (Р—, Ф+) признак, свойственный иностранному языку, не сразу вносится в третью систему (например, различение притяжательного детерминатива и местоимения, прошедшего и предшествования прошедшему). На начальном этапе языковой контакт приводит к стиранию разделительной границы между членами оппозиции в иностранном языке и распространению значений одного из членов оппозиции на область значений другого члена. Таким образом, на какой-то период третья система оказывается лишенной дифференциального признака иностранного языка.

б) При наличии дифференциального признака в родном языке и отсутствии данного признака в иностранном языке (Р+, Ф—) в третью систему вносится признак, свойственный родному языку (например, противопоставление личного и возвратного поссесива, видовой признак в системе будущего времени). Языковой контакт приводит к появлению оппозиции, несвойственной иностранному языку. Таким образом, третья система приобретает лишний дифференциальный признак, не присущий данному иностранному языку.

Если в родном языке есть дифференциальный признак, несвойственный иностранному языку (Р+, Ф—), но он не подкрепляется каким-либо другим признаком, по которому также различались бы члены данной оппозиции (и на который иностранный язык мог бы реагировать), как мы это видели на примере plus-que-parfait и русского прошедшего совершенного и несовершенного вида, то интерференция создается только потенциально. Такой признак не вносится в третью систему, интерференция не реализуется.

в) Наличие однородного дифференциального признака в обоих языках (Р+, Ф+), при различной дистрибуции членов оппозиции, ведет к перенесению данного признака в третью систему с дистрибуцией, свойственной родному языку (например, passé composé — imparfait и русское прошедшее совершенного и несовершенного вида). Разделительная граница между членами оппозиции в иностранном языке перемещается, принимая положение, свойственное разделительной границе в родном языке, что ведет к изменению дистрибуции между членами оппозиции.

ON THE PROBLEM OF GRAMMATICAL INTERFERENCE

Summary

The authors analyse the process and types of grammatical interference. By grammatical interference they mean a deviation from the norm of the language L under the influence of the language L_1 which takes place owing to the contact between the two languages. The process of interference consists in replacing the system of grammatical distinctive features of L by a different system which grows up in the mind of the speaker under the influence

of the system of grammatical distinctive features of L_1 (his mother tongue). The third system constructed in that way contains distinctive features of both languages. The authors analyse Russian-French interference facts and single out the following three types of grammatical interference: 1) no distinctive features of L are introduced into the third system; 2) some features of L_1 not characteristic of L are introduced into the third system; 3) there is a change in distribution of those distinctive features of L which are also characteristic of L_1, but have in it a different correlation of content and expression.

---

www.ruslang.ru

II. ФОНОЛОГИЯ

С. К. Шаумян

ПАНХРОНИЧЕСКАЯ СИСТЕМА ДИФФЕРЕНЦИАЛЬНЫХ ЭЛЕМЕНТОВ И ДВУХСТУПЕНЧАТАЯ ТЕОРИЯ ФОНОЛОГИИ

В структурно-типологическом плане большой интерес представляет панхроническая система дифференциальных элементов, предложенная Р. Якобсоном, Г. Фантом и М. Халле. В эту систему входят следующие 12 бинарных дифференциальных элементов: 1) вокальность — невокальность, 2) консонантность—неконсонантность, 3) прерывность — непрерывность, 4) абруптивность — неабруптивность, 5) яркость — тусклость, 6) звонкость—глухость, 7) компактность — диффузность, 8) низкая тональность — высокая тональность, 9) бемольная тональность — простая тональность, 10) диезная тональность — простая тональность, 11) напряженность — ненапряженность, 12) назальность — неназальность. Эти дифференциальные элементы образуют в совокупности универсальную типологическую схему, пригодную для определения системы дифференциальных элементов в любом языке мира.

Данная панхроническая система дифференциальных элементов в силу своей общности и абстрактности должна, естественно, рассматриваться в качестве рабочей гипотезы, которая нуждается в проверке по двум направлениям: 1) в концептуальном плане, с точки зрения логической обоснованности фундаментальных понятий этой гипотезы, и 2) в эмпирическом плане, с точки зрения действительной применимости этой гипотезы к описанию фонологических систем всех конкретных языков мира.

В настоящей статье мы выскажем некоторые соображения по поводу концептуальной стороны панхронической системы дифференциальных элементов в связи с предлагаемой нами двухступенчатой теорией фонологии.

Если анализировать рассматриваемую панхроническую систему дифференциальных элементов в концептуальном плане, то прежде всего возникает следующая фундаментальная проблема: каков точный смысл понятий «дифференциальный элемент» и «избыточный элемент», на которые опирается панхроническая система дифференциальных элементов?

Данная проблема неизбежно возникает в связи с методом разграничения дифференциальных и избыточных элементов, применяемым авторами рассматриваемой системы. Так, опираясь на факт, что напряженность и аспирация никогда не противопоставляются друг другу в одних и тех же фонологических условиях (например, если в одном языке различительной функцией обладает напряженность, то в другом языке в соответствующей роли выступает аспирация, или же в одном и том же языке в одних позициях выступает ненапряженность, а в других позициях ей соответствует аспирация), авторы сводят противопоставления «напряженность—ненапряженность» и «аспирация—неаспирация» к одному противопостав-

лению: «напряженность—ненапряженность». Таким образом, противопоставление «аспирация—неаспирация» считается избыточным. Аналогичным образом, сводятся друг к другу другие противопоставления. Если рассматривать дифференциальные элементы как физические противопоставления, то сведение противопоставления «аспирация—неаспирация» к противопоставлению «напряженность—ненапряженность» представляется произвольным: в самом деле, почему первое противопоставление должно сводиться ко второму, а не наоборот — второе к первому?

Выход из этой трудности и ряда аналогичных трудностей, связанных с проблемой тождества фонологических противопоставлений, позволяет найти предлагаемая нами двухступенчатая теория фонологии, изложенная подробно в специальной статье [1].

В двухступенчатой теории фонологии мы расщепляем понятие дифференциального элемента на понятия дифферентора и дифферентоида, а понятие фонемы — на понятия фонемы и фонемоида. В дифференторе нет ничего физического, это чисто реляционный элемент (так называемый конструкт), субстратом которого служит реляционно-физический элемент— дифферентоид. Фонема должна считаться пучком дифференторов. Если в дифференторах нет ничего физического, то и в фонеме не может быть ничего физического. При таком понимании фонема — это тоже конструкт, субстратом которого служит реляционно-физический элемент — фонемоид. В рамках современной логики науки мы относим дифферентоиды и фонемоиды, с одной стороны, и дифференторы и фонемы, с другой, — к двум принципиально отличным друг от друга ступеням абстракции: к ступени наблюдения и к ступени конструктов. Дифферентоиды и фонемоиды относятся к ступеням наблюдения, а дифференторы и фонемы — к ступени конструктов. Дифферентоиды и фонемоиды в качестве элементов субстанции есть переменные, которые в зависимости от обстоятельств могут быть акустическими, графическими, тактильными или цветовыми элементами. Что же касается дифференторов и фонем, то это постоянные элементы, природа которых остается одинаковой независимо от их конкретных физических субстратов.

Такова вкратце сущность двухступенчатой теории фонологии. Следует подчеркнуть, что в двухступенчатой теории фонологии строго различаются следующие три рода элементов: 1) звуки и акустические элементы звуков, 2) фонемоиды и дифферентоиды, 3) фонемы и дифференторы. Первые — чисто физические элементы, вторые — реляционно-физические элементы, третьи — чисто реляционные элементы (конструкты).

Если рассматривать приведенные выше противопоставления с точки зрения двухступенчатой теории фонологии, то мы в праве отождествлять напряженность и аспирацию только в том случае, если будем считать их физическими субстратами тождественных дифференторов.

Обозначив отношение воплощения знаком I, напряженность — знаком T и дифферентор, воплощаемый напряженностью, — D_1, мы можем представить связь напряженности со своим дифферентором в виде следующей формулы:

$$I\ (T,\ D_1)$$

Обозначив аспирацию знаком A и дифферентор, воплощаемый аспирацией, знаком D_2, мы можем представить связь аспирации со своим дифферентором в виде формулы:

$$I\ (A,\ D_2)$$

[1] См.: С. К. Шаумян. Двухступенчатая теория фонемы и дифференциальных элементов. «Вопросы языкознания», 1960, № 5; его же. Преобразование информации в процессе познания и двухступенчатая теория структурной лингвистики (в настоящем сборнике, стр. 5—12).

Поскольку D_1 и D_2 тождественны друг другу, то отсюда следует и тождество T и A как субстратов D_1 и D_2.

Возникает терминологический вопрос: как назвать дифференторы D_1 и D_2? Используя терминологию, применяемую авторами панхронической системы дифференциальных элементов, мы можем, если угодно, назвать D_1 и D_2 разновидностями напряженности, но в таком случае мы должны строго различать и не смешивать друг с другом термин «напряженность» как конвенциональное обозначение определенного дифферентора и термин «напряженность» как обозначение реляционно-физического свойства, служащего субстратом дифферентора. Принимая данное терминологическое условие, мы можем утверждать, что напряженность-дифферентор воплощается двумя дифферентоидами: напряженностью и аспирацией.

Применяя аналогичные рассуждения к остальным терминам, мы должны различать звонкость-дифферентор и звонкость-дифферентоид, компактность-дифферентор и компактность-дифферентоид, абруптивность-дифферентор и абруптивность-дифферентоид и т. д. Чтобы не смешивать дифференторы и дифферентоиды, можно ввести индексацию: в тех случаях, когда названия дифференторов и дифферентоидов совпадают, мы будем обозначать индексом 1 дифферентоиды и индексом 2 — дифференторы, например, звонкость$_1$ и звонкость$_2$, компактность$_1$ и компактность$_2$, абруптивность$_1$ и абруптивность$_2$ и т. д.

В свете двухступенчатой теории фонологии мы должны расщепить понятие панхронической системы дифференциальных элементов на два понятия — понятие панхронической системы дифференторов и понятие физического воплощения панхронической системы дифференторов.

В приведенную выше систему 12 бинарных противопоставлений в качестве дифференторов представляется необходимым внести следующие изменения. Вокальность$_2$ — невокальность$_2$ и консонантность$_2$ — неконсонантность$_2$ сводятся к одному противопоставлению: вокальность$_2$ — консонантность$_2$. Напротив, противопоставление компактность$_2$ — диффузность$_2$ целесообразно расщепить на два противопоставления: компактность$_2$ — некомпактность$_2$, диффузность$_2$ — недиффузность$_2$. Кроме того, необходимо ввести в систему следующие противопоставления: плавность$_2$ — неплавность$_2$, скольжение$_2$ — нескольжение$_2$, долгота$_2$ — краткость$_2$.

Остановлюсь теперь на обосновании указанных изменений.

Прежде всего следует остановиться на необходимости введения двух особых дифференторов — плавности$_2$ и скольжения$_2$.

К плавным фонемам относятся такие, как *r* или *l*, а к скользящим фонемам — такие, как *j* или *h*. Р. Якобсон и его соавторы сводят плавность к соединению вокальности и консонантности, а скольжение — к соединению невокальности и неконсонантности.

Спрашивается: могут ли существовать фонемы, которые являются одновременно и гласными, и согласными, и, наоборот, фонемы, которые не являются ни гласными, ни согласными?

Ответ на этот вопрос зависит от того, как трактовать природу фонемы.

Если считать фонему акустическим элементом, то, конечно, можно допустить, что существуют фонемы, соединяющие в себе физические свойства гласных и согласных, или, напротив, не обладающие физическими свойствами ни тех, ни других. Но как уже сказано, фонема не заключает в себе ничего физического.

Если считать фонему конструктом, то ответ на поставленный вопрос должен быть отрицательным. Суть дела в том, что гласные и согласные фонемы, рассматриваемые как конструкты, не могут противопоставляться в одной и той же позиции. Гласные и согласные фонемы как конструкты взаимно исключают друг друга в одной и той же позиции. Поэтому вокальность и консонантность есть синтагматические, а не парадигмати-

ческие дифференциальные элементы[2]. Отсюда следует, что не могут существовать фонемы, соединяющие в себе вокальность и консонантность. С другой стороны, не может быть фонем, не являющихся ни гласными, ни согласными; всякая фонема как конструкт должна быть либо гласной, либо согласной.

У Р. Якобсона и его соавторов встречаются примеры, которые на первый взгляд противоречат только что сказанному. Так, из английского языка приводится противопоставление *yet* [i'et] — *vet*, *set*, *net* в доказательство того, что гласные и согласные могут противопоставляться в одной и той же позиции [3]. Но в действительности во всех подобного рода примерах гласные и согласные фонемы находятся не в одинаковых, а в разных позициях. Суть дела в том, что понятие позиции также нуждается в расщеплении на два понятия: на понятие наблюдаемой позиции и понятие позиции-конструкта. В необходимости этого расщепления легко убедиться при анализе односложных слов. Если бы в словах, скажем, *дар* и *акт* имелись противопоставления в начале слов *д—а*, и в середине *а—к*, то посредством подстановки согласного вместо гласного и наоборот мы могли бы изучить сочетание *dkt* вместо *akt* и сочетание *dkr* вместо *dar*. Но такие сочетания принципиально невозможны в русском языке. На самом деле противопоставление *дар—акт* необходимо представить в виде схемы:

dar0
0akt

В этой схеме согласные фонемы-конструкты *д* и *т* противопоставляются нулям согласных (нуль в данном случае тоже есть конструкт).

Аналогичным образом, противопоставление *yet* и *set* в приведенном выше примере следует представить в виде схемы

i0et
0set

В этой схеме гласная фонема-конструкт противопоставляется нулю гласных, а согласная фонема-конструкт противопоставляется нулю согласных.

Из всего сказанного с неизбежностью следует необходимость свести две пары дифференторов вокальность$_2$ — невокальность$_2$ и консонантность$_2$ — неконсонантность$_2$ к одной паре вокальность$_2$ — консонантность$_2$. Вместе с тем становится ясным, почему нужны новые самостоятельные дифференторы: плавность$_2$ — неплавность$_2$ и скольжение$_2$ — нескольжение$_2$.

Что касается долготы и краткости, то, рассматривая долготу с физической точки зрения, Р. Якобсон и его соавторы относят ее к просодическим элементам. В действительности же, если рассматривать долготу как конструкт, то между долгими и краткими фонемами нет никакой количественной разницы. Количественная разница существует только между соответствующими фонемоидами. С точки зрения ступени конструктов, долгота — это не количество, а качество. Отсюда становится ясной необходимость рассматривать долготу и краткость как дифференциальные элементы, равноправные остальным дифференциальным элементам.

---

[2] При более строгом рассмотрении сущности вокальности и консонантности можно доказать, что вокальность и консонантность — это вообще не дифференциальные элементы, а элементы, родственные просодическим элементам. Оставляя исследование этого сложного вопроса для другого случая, мы следуем в настоящей статье общепринятой точке зрения, трактующей вокальность и консонантность в качестве дифференциальных элементов.

[3] R. Jakobson, G. Fant, M. Halle. Preliminaries to speech analysis. Cambridge, Mass., 1952, стр. 20.

Противопоставление компактность$_2$ — диффузность$_2$ целесообразно расщепить на противопоставления компактность$_2$ — некомпактность$_2$ и диффузность$_2$ — недиффузность$_2$ с целью последовательного проведения принципа бинарности (во избежание комплексных дифференторов со знаками +, —). Этот прием является вполне правомерным на ступени конструктов.

В итоге мы предлагаем панхроническую систему дифференторов, в которую входят следующие 15 бинарных противопоставлений: 1) вокальность$_2$ — невокальность$_2$, 2) компактность$_2$ — некомпактность$_2$, 3) диффузность$_2$ — недиффузность$_2$, 4) низкая тональность$_2$ — высокая тональность$_2$, 5) бемольная тональность$_2$ — простая тональность$_2$, 6) назальность$_2$ — неназальность$_2$, 7) непрерывность$_2$ — прерывность$_2$, 8) абруптивность$_2$ — неабруптивность$_2$, 9) яркость$_2$ — неяркость$_2$, 10) звонкость$_2$ — незвонкость$_2$, 11) диезная тональность$_2$ — простая тональность$_2$, 12) напряженность$_2$ — ненапряженность$_2$, 13) плавность$_2$ — неплавность$_2$, 14) скольжение$_2$ — нескольжение$_2$, 15) долгота$_2$ — краткость$_2$.

Данную панхроническую систему дифференторов нужно рассматривать в качестве рабочей гипотезы, которая, несомненно, подвергнется новым модификациям в результате дальнейшего концептуального анализа и углубления в эмпирический материал.

PANCHRONISTIC SYSTEM OF DISTINCTIVE FEATURES

Summary

From the view-point of the two-level theory of phonology developed by the author the article deals with the panchronistic system of distinctive features as presented by R. Jakobson, G. Fant and M. Halle. Distinguishing between the observation level which encompasses differentoids and phonemoids, and the construct level which includes differentors and phonemes, the author considers it necessary to draw the same line of distinction in the above panchronistic system of distinctive features. As a result of investigation he suggests a panchronistic system of differentors consisting of 15 binary oppositions.

И. И. Ревзин

ОБ ОДНОМ ПОДХОДЕ К МОДЕЛЯМ ДИСТРИБУТИВНОГО ФОНОЛОГИЧЕСКОГО АНАЛИЗА

§ 1. **Идея моделирования и обычные определения фонемы.** Идея моделирования играет все более важную роль в современной лингвистике. Моделирование необходимо предполагает упрощение основных понятий и уточнение исходной ситуации. Нам представляется, что при создании фонологических моделей обычное определение фонемы не может быть положено в основу, так как оно прибегает к понятию значения, которое пока формализовать не удалось. В самом деле, если отвлечься от всяких терминологических тонкостей, то обычное определение фонемы сводится к следующему. Два звука принадлежат разным фонемам, если имеется хотя бы одна пара слов, различающихся только этими звуками и имеющих разное лексическое значение. По этому определению [ɛ] и [e] в немецком языке две фонемы (есть пара слов *Bett*, *Beet*), в то время как [ɛ] — «открытое» в русском, например в *этот* и [e] — «закрытое», например, в *эти* — принадлежат одной фонеме (нельзя найти пары разных по значению слов, где бы эти звуки противопоставлялись). Это определение имеет, однако, два крупных недостатка:

1) Согласно этому определению, гласный, например [у], и согласный, например [ф], могут оказаться принадлежащими одной фонеме (если не найдется пары слов, где бы эти звуки явились единственными смыслоразличителями). Правда, часто вводится оговорка, что варианты одной фонемы должны быть физически близки. Но тогда в определение приходится вводить разнородные элементы.

2) Это определение неудобно потому, что заставляет обращаться к лексическому значению слов, а здесь теоретически неясно, что значит утверждение «два слова имеют разное значение». В самом деле, если человек говорит: *слава богу* — с фрикативным [γ], но *бог Аполлон* или *бог света* — с [г], соответственно — с [к], то трудно установить, имеется ли здесь два слова с разными значениями, или одно слово с одним значением.

Конечно, в структурной лингвистике на самом деле обращаются не к значению данного слова, а, так сказать, к номеру его значения (в этом суть теста коммутации). Тем не менее и при этом не снимаются трудности, мешающие последовательной формализации.

§ 2. **Принципы дескриптивной фонологии.** В работах американских лингвистов [1] был впервые выдвинут тезис, что значение не имеет никакого отношения к определению фонемы и что можно установить фонемати-

[1] См.: Ch. H o c k é t t. A system of descriptive phonology. «Language», vol. XVIII, № 1, 1942; B. B l o c h. A set of postulates for phonemic analysis. «Language», vol. XXIV, № 1, 1948, стр. 6—9.

ческую систему языка, не зная значения слов, только на основании анализа распределения.

Возможность чисто дистрибутивного анализа системы фонем некоторого языка была ясна уже создателю современной фонологии Н. С. Трубецкому, который приводит следующий пример чисто дистрибутивного анализа [2]. Возьмем аттический диалект древнегреческого языка. В нем есть один звук, встречающийся только в начале слова, а именно spiritus asper. Возьмем все звуки, которые стоят за spiritus asper, назовем этот класс гласными, а остальные звуки согласными и так далее. Таким методом Трубецкой и строит полную классификацию фонем данного диалекта.

Однако то, что у Трубецкого было некоторым вспомогательным приемом исследования, стало в дескриптивной лингвистике основным методом.

Само деление речевого континуума на отрезки (сегменты) должно, согласно этой концепции, происходить на основе анализа распределения [3], этот же принцип должен быть положен в основу объединения различных сегментов в классы. Обращение к носителю языка допускается лишь для того, чтобы определить так называемые свободные варианты (free variants) одной и той же фонемы (ср. разные произношения последнего звука в русском слове *дождь*, которые не зависят от позиции). Так называемые «связанные варианты» (bound variants) должны устанавливаться целиком на основе анализа окружения.

Дескриптивной лингвистикой разработан ряд приемов для проведения этой работы, однако вся процедура является исключительно громоздкой. Для дальнейшей формализации необходимо эту ситуацию упростить и уточнить.

§ 3. **Дескриптивная фонология и спектральный анализ.** Дескриптивная фонология до сих пор опиралась на слуховой анализ потока речи, что не дает достаточно объективных исходных данных. Современная фонетика располагает вполне объективным материалом, основанном на спектральном анализе звуков речи. Самое важное достижение спектрального анализа состоит в том, что такие признаки звуков, давно выделенные в фонетике, как «гласность — согласность», «глухость — звонкость», «открытость — закрытость» и т. д., являются вполне объективными и имеют соответствующие аналоги на языке спектрограмм.

Так, Р. Якобсон выделил на основе анализа спектрограмм 12 противопоставлений, которые, как он считает, применимы к анализу всех языков мира [4].

Наша задача состоит теперь в том, чтобы построить дистрибутивную фонологическую модель на основании уточнения и упрощения тех данных, которые предоставляет в наше распоряжение спектральный анализ. Иначе говоря, мы хотим точно сформулировать те предположения, на которых построена наша фонологическая модель (вернее, фонологические модели, поскольку будут построены две модели).

1) Прежде всего мы формулируем предположение, что путем спектрального анализа может быть произведена естественная сегментация звукового континуума. Ясно, что это предположение весьма существенно упрощает действительное положение, но здесь можно было бы сослаться на мнение такого видного специалиста в области акустической фонетики, как Г. Фант, который считает, что «имеется естественная сегментация речевого потока на звуки и части звуков» [5].

---

[2] Н. С. Трубецкой. Основы фонологии. Пер. с нем. М., 1960, стр. 272—273.

[3] См.: Z. Harris. Methods in structural linguistics. Chicago, стр. 35 и сл.

[4] См., например: R. Jakobson, G. Fant, M. Halle. Preliminaries to speech analysis. «MJT Technical Report», № 13, 1955, стр. 40.

[5] См.: G. Fant. Acoustic theory of speech production. «Royal Institute of Technology, Division of Telegraphy-Telephony». Report, № 10, 1958, стр. 17.

2) Второе предположение состоит в том, что каждый звук может быть представлен в виде совокупности n признаков, или, как говорят, n-ки отдельных признаков.

3) Мы будем исходить из предпосылки, что фонемы суть единицы определенного кода, в то время как звуки есть единицы сообщения, протекающего в определенных условиях, т. е. подверженного искажению под влиянием помех.

Задача фонологического анализа, как он понимается здесь, есть задача обнаружения кода по данным сообщениям. Для этого постараемся выяснить, каковы в основном те помехи, которые могут «искажать» сообщение.

Во-первых, это индивидуальные особенности произношения говорящих, зависящие от целого ряда причин (например, пол, возраст и т. п.). Мы будем для простоты считать, что искажения этого типа нам удалось устранить. Этим подчеркивается, что модель прилагается к некоторой идеализированной ситуации, где не потребуется иных критериев, кроме критерия дистрибуции.

Другим, и в сущности, основным искажением, с которым имеет дело фонология, является то обстоятельство, что для каждого звука «шумом», «помехой» может оказаться соседний с ним звук. Речь идет о так называемом комбинаторном изменении звуков, которое приводит к состоянию дополнительного распределения звуков речи.

Как формализовать эту ситуацию?

Мы видели, что путем анализа спектрограмм каждый звук может быть представлен в виде n-ки признаков. Кроме того, все позиции звуков внутри слова можно перенумеровать (первый звук от конца, второй от конца и т. п.).

Пусть какой-нибудь признак x_i обязательно встречается в данной n-ке признаков, если в предшествующей n-ке был признак x_k или в последующей n-ке был признак x_l. Тогда данный признак x_i называется связанным по отношению к x_k или x_l в данной позиции. Отбросим в каждой n-ке все признаки, связанные в данной позиции по отношению к некоторым другим признакам. Оставшиеся признаки назовем существенными («релевантными») для данной позиции данной n-ки. Признак x_i называется дифференциальным признаком n-ки n_1 по отношению к n-ке n_2 в данной позиции, если он существенен для n_1 и отсутствует в n_2 в данной позиции.

Предложенную теорию можно обобщить следующим образом. Так как пробел между словами в теории кодов рассматривается как особый звук, мы будем считать, что между каждыми двумя словами находится «пустой звук», т. е. n-ка, состоящая из «пустых признаков». Тогда связанным по отношению к этой n-ке окажется, например, в русском языке признак глухости в конце слова. Подобное явление называется в фонологии нейтрализацией.

Мы говорили о признаках, дифференциальных по отношению к данной позиции. Собрав признаки дифференциальные для позиции данной n-ки, мы получим ее общие дифференциальные признаки [6].

Мы видим, что помехи, связанные с действием соседних звуков, могут быть устранены объективным методом анализа. Теперь мы можем по сообщению восстановить единицу кода, а именно фонему.

Фонема есть совокупность (или, как иногда говорят, пучок) общих дифференциальных признаков. Две разные n-ки суть манифестации

---

[6] При таком подходе к выделению дифференциальных признаков, по-видимому, отпадут некоторые обоснованные критические замечания проф. П. С. Кузнецова по адресу теории Якобсона (см.: П. С. Кузнецов. О дифференциальных признаках фонем. «Вопросы языкознания», 1958, № 1, стр. 55—56), в частности — замечание о том, что «целые категории звуков данного языка могут отличаться друг от друга одновременно двумя признаками» (стр. 57).

одной фонемы, если их общие дифференциальные признаки совпадают. В противном случае они суть манифестации разных фонем.

Рассмотрим теперь ряд примеров, обычно приводимых при обсуждении проблемы фонемы, с точки зрения этих определений.

У [е] и [ɛ] в русском языке признак закрытости связан с признаком мягкости в следующем согласном. Это верно для любой позиции]е] в слове. Отсюда общие дифференциальные признаки для п-ки [е], т. е. для совокупности всех физических качеств этого звука, в русском языке те же, что и общие дифференциальные признаки для п-ки [ɛ], хотя эти п-ки (т. е. звуки) сами по себе различны. Для немецкого языка признак закрытости не связан с каким-нибудь признаком предшествующего или последующего согласного.

Возьмем русские [ы] и [и]. Признак заднего образования в [ы] связан с признаком твердости предшествующего согласного и поэтому связан во всех позициях. Он не является общим дифференциальным признаком, и поэтому общие дифференциальные признаки для [ы] и [и] в русском языке совпадают. Отсюда, по нашей теории, [и] и [ы] — два варианта одной фонемы.

Рассмотрим теперь п-ки признаков, характеризующих звуки [д] и [т] в русском языке. Признак глухости является связанным в конце слова, поэтому он не является дифференциальным для этой позиции. Однако глухость является общим дифференциальным признаком для [д]. Отсюда [д] и [т] в русском языке — разные фонемы. Что же касается позиции в конце слова, то здесь происходит нейтрализация. В корейском же языке употребление *д* или *т* зависит от места в слове, в начале слова встречается только *т*, а в середине слова между гласными только *д* [7]. Поэтому глухость связана здесь во всех позициях и не является дифференциальным признаком.

Отличие изложенной теории от многих абстрактных фонологических теорий в том, что она связывает акустическое и фонологическое изучение звуков речи. Акустические данные являются тем исходным материалом, с которого начинается фонологический анализ. Задача акустического (спектрального) анализа состоит в разложении кривой на ряд дискретных признаков. Задача фонологического анализа состоит в выделении из этих признаков общих дифференциальных признаков и в идентификации фонем.

§ 4. **Фонема в московской фонологической школе и ее формализация.** Существует еще одна важная единица фонологической системы, которая рассматривается московской фонологической школой и также называется в ней фонемой. Это та единица, которая остается тождественной в разных алломорфах одной и той же морфемы. Например, в словах *вод* [вод] и *вода* [вада] два алломорфа одной морфемы, поэтому [о] и [а] считаются представителями одной «фонемы» (в этом понимании) или, как говорит Р. И. Аванесов, одного «фонемного ряда» [8]. Этим последним термином мы и будем пользоваться в дальнейшем. Полностью формализовать это понятие в дистрибутивных терминах можно будет, по-видимому, лишь тогда, когда будет построена полная система абстрактной морфологии. Можно, однако, попытаться найти хотя бы частичный дистрибутивный аналог этого важного понятия.

Прежде всего, расширим исходные предпосылки. Мы будем считать, что, кроме исходных предположений § 3, нам задано разбиение всех звуков на гласные и согласные, в каждом слове отмечено место ударения и в понятие фонетической позиции каждого гласного входит расстояние от ударного гласного в числе слогов.

[7] Л. Р. Зиндер. Специфические особенности восприятия звуков речи. Восприятие звуковых сигналов в различных акустических условиях. М., 1959, стр. 70.

[8] Р. И. Аванесов. Фонетика современного русского литературного языка. М., 1956, стр. 31.

Кроме этого, рассмотрим одно понятие, которое может быть сформулировано в терминах § 3. Мы будем говорить, что две фонемы F_1 и F_2 находятся в отношении корреляции по признаку x_i, если они различаются только по этому признаку. Мы будем говорить, что две фонемы входят в одну элементарную фонологическую категорию, если они находятся в отношении корреляции по признаку x_i и x_j имеет хотя бы одну позицию нейтрализации.

Попытаемся теперь обобщить понятие элементарной фонологической категории. Для этого введем следующие вспомогательные определения.

Если в некоторой фонетической позиции фонема (или вариант фонемы) F_1 встречается, а фонема (или вариант фонемы) F_2 не встречается (причем F_1 и F_2 — обе гласные или обе согласные), то мы будем говорить, что F_1 доминирует над F_2 в данной позиции.

Например, после твердого согласного в первом предударном слоге встречается [а], но не встречается [о]. Значит, в этой позиции [а] доминирует над [о]. Другой пример: [к] встречается в конце слова (перед «нулевой» фонемой), а [г] не встречается. В этой позиции [к] доминирует над [г].

Мы будем говорить, что фонема F_1 абсолютно доминирует над фонемой F_2 в данной позиции, если для всякой другой фонемы F, которая доминирует над F_2 в данной позиции, найдется фонема F_3 такая, что F доминирует над F_3, причем количество признаков, различающих F и F_3, меньше, чем количество признаков, различающих F и F_2.

Рассмотрим теперь случай, когда над F_0 доминирует несколько фонем F_1, F_2, . . ., F_K. Из наших определений логически не следует, что одна из них должна абсолютно доминировать над F_0. Между тем, как будет видно из примеров, в реальных языках, например в русском языке, мы, по-видимому, имеем дело именно с такой ситуацией. Поэтому введем следующее дополнительное определение. Мы будем говорить, что система фонем фонологически правильна, если в ней для всякой фонемы F_0 из условия «F_1, F_2, . . ., F_K — *все фонемы, доминирующие над* F_0», следует, что хотя бы одна фонема F_i абсолютно доминирует над F_0.

В дальнейшем рассматриваются только фонологически правильные системы.

Если F_1 абсолютно доминирует над F_2, то мы будем говорить, что F_1 и F_2 входят в один фонемный ряд.

Примеры. Легко видеть, что в конце слова (перед нулевой фонемой) [п] доминирует над [г]. В самом деле [п] встречается в конце слова, а [г] не встречается в конце слова. Однако, [п] и [г] не входят в один фонемный ряд, так как [п] не абсолютно доминирует над [г]. В то же время [к] абсолютно доминирует над [г], так как, например, для [п], доминирующей над [г], найдется [б], над которым доминирует [п], причем количество признаков, различающих [п] и [б], меньше, чем число признаков, различающих [п] и [г]. В этом примере мы не получили ничего нового; фонемный ряд просто совпадает с элементарной фонологической категорией (конец слова — позиция нейтрализации для признака звонкости согласного).

Но этот же принцип можно, по-видимому, применить и к анализу русских гласных, и на этой основе построить модель, отражающую сложные взаимоотношения фонем в русском языке. В самом деле [и^{е}][9] доминирует над [о] в позиции предударного слога, но мы не объединим [и^{е}] и [о] в один фонемный ряд. В то же время [а] и [о] объединяются в один фонемный ряд, так как есть фонема [е], как, например, в *кедр*, такая, что [и^{е}] доминирует над [е], причем разница между [и^{е}] и [е] меньше, чем разница между [о] и [и^{е}].

[9] В транскрипции Р. И. Аванесова, напр. *кедровый* — [киедровъ̮].

Мы видели, что для случая [к] и [г] фонемный ряд совпал с элементарной фонологической категорией. Каково же отношение между этими понятиями в общем случае?

Покажем, что всякие две фонемы, входящие в одну элементарную категорию, входят и в один фонемный ряд, но обратное неверно.

В самом деле, две фонемы F_1 и F_2, входящие в одну элементарную фонологическую категорию, по определению различаются только в одном признаке x_i, который принимает разные значения для F_1 и F_2, причем x_i связан в некоторой позиции P. Пусть для определенности он принимает значение a этой позиции и пусть это то значение, которое этот признак принимает для F_1. Тогда F_2 никогда не встречается в этой позиции, а значит, F_1 доминирует над F_2.

Покажем теперь, что F_1 абсолютно доминирует над F_2. Предположим противное, а именно, что F_1 не абсолютно доминирует над F_2. Поскольку система фонологически правильна, то имеется F_0, абсолютно доминирующая над F_2. Это означает наличие F_0 такой, что F_0 доминирует над F_2 и количество признаков, по которым различаются F_0 и F_2, меньше, чем количество признаков, по которым различаются F_1 и F_2. Поскольку F_1 и F_2 различаются только по одному признаку, то F_0 и F_2 совпадают. Но этого, разумеется, быть не может. Полученное противоречие доказывает, что F_1 абсолютно доминирует над F_2.

Таким образом, всякая элементарная фонологическая категория есть фонемный ряд. С другой стороны, построим фонемный ряд так, чтобы в него входили две фонемы, различающиеся двумя и более признаками. Такой фонемный ряд не может быть элементарной фонологической категорией. Наше утверждение доказано.

Две последовательности фонем считаются *морфологически тождественными*, если на соответственных местах стоят одинаковые фонемы или же — в слабых позициях — фонемы, входящие в один фонемный ряд. (Мы называем позицию слабой, если в ней встречаются не все фонемы какого-нибудь фонемного ряда. В частности, всякая позиция нейтрализации — слабая.)

В заключение отметим следующее. Очевидно, что два звука речи, относимые к разным фонемам по смыслоразличительному критерию, будут отнесены к разным фонемам и разным фонемным рядам в построенных моделях, так как наличие пары, различающейся по смыслу, говорит о релевантности признака в данной позиции. Обратное же очевидно неверно. В связи с этим дистрибутивные модели являются полезным, но все же вспомогательным инструментом фонологического исследования.

AN APPROACH TO THE MODELS OF DISTRIBUTIONAL PHONOLOGICAL ANALYSIS

Summary

In an abstract phonological model the «phoneme» may be defined in purely distributional terms as a bundle of non-bound features (free variants being eliminated statistically).

The author develops a tentative descriptive model for what is called the «phoneme» in the Moscow phonological school. Some natural generalization of the concept of «neutralization» is found which can be applied to the Russian vowels in the non-stressed position.

Э. Василиу

ФОНОЛОГИЧЕСКОЕ ОПИСАНИЕ РУМЫНСКОГО ВОКАЛИЗМА

0. Описание, которое я собираюсь здесь сделать, отличается от сделанных до настоящего времени (Граур-Росетти, Петрович, Росетти, Малецкий, Андрей Аврам). Это отличие касается в первую очередь методологии.

0. 1. Я избрал отправной точкой ясное различие между формой (фонологической) и субстанцией (фонической). То есть, редукция вариантов так же, как классификации, основана не на фонетических критериях, а на функциональных критериях (парадигматических и синтагматических). Иными словами, я пытался придать фонологической реальности интралингвистическое описание, являющееся выводом из «имманентной структуры» языка (Шаумян). Фонетический элемент рассматривается не как критерий классификации, а как элемент, который сам должен быть классифицирован.

0. 2. В то же время следует отметить, что, по моему мнению, возможность дать чисто формальное описание фонологической реальности, то есть описание, которое никогда не сталкивалось бы на каком-либо этапе с фонической субстанцией, является чистой иллюзией. Всякое фонологическое описание до настоящего времени (включая сюда глоссематические описания) рассматривает так или иначе «фонетическую близость» как средство идентификации фонологических единиц.

С. К. Шаумяну принадлежит заслуга создания формального статуса этих «сталкиваний» с фонетической субстанцией.

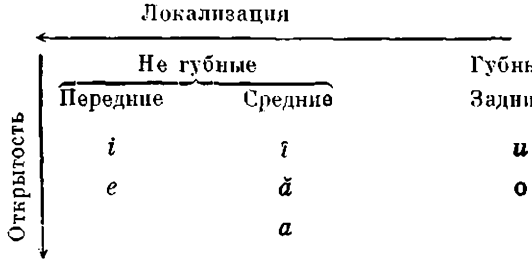
а) его теория фонем рассматривает фоническую субстанцию как член отношения между «конструктом» (фонемой) и «непосредственной реальностью» (звуком языка);

б) «фоническая близость» сформирована как «законы трансформации».

0. 3. Процесс описания, который я применяю, следующий: а) испытание коммутации, б) отношение элементов, установленных коммутацией, к фонической субстанции, в) правила трансформации, г) редукция вариантов, д) фонологическая система.

1. Рассмотрим следующие пары: [der — dor — dur], [mări — meri], [vînă — vinǎ]. Здесь можно установить элементы [a, o, u, ă, î, e, i] как элементы, различающиеся между собой.

1. 1. Эти элементы представляют звуки, фонетическое описание которых исчерпывается, если построить треугольник гласных.



1.2. Теперь рассмотрим следующие пары: [bẹată — bjată], [sọare — sare], [lup^{j} — lupi], [lew — le^{w}u], [le^{j} — le^{j}i].

Здесь можно установить следующие единицы [ẹ, ọ, j, w, ^{w}u, ^{j}i].

1.2.1. Не входя в подробности, можно дать этим элементам следующее *фонетическое описание*: [ẹ, ọ, j, w] *являются асиллабическими* [e, o, i, u]; [^{w}u, ^{j}i] являются [u, i], которые начинаются с фрикции того же тембра гласной; [i̯] «псевдо-конечное [i]» (Петрович) является асиллабическим и безгласным [i].

1.2.2. Ряд [a, e, i, o, u, ă, î, ẹ, ọ, j, w, i̯, ^{j}i,^{w}u] является максимальным рядом, который можно установить коммутацией. Аллофоны, которые будут описаны в пункте 2, могут быть установлены или коммутацией, или правилами трансформации. Мы предпочитаем последний путь, поскольку он наиболее удобен.

2. Можно сформулировать следующие правила трансформации, где «оператор трансформации» является «позицией» (ср. Шаумян).

2.1. После перерыва (##) так же, как после гласных, ряд [e, i, o, u] представлен рядом [^{j}e, ^{j}i, ^{w}o, ^{w}u] или же рядом [e, i, o, u]: [^{j}elefant ~ elefant, ^{j}inginer ~ inginer, ^{w}operă ~ operă, ^{w}universitate ~ universitate] или [ale^{j}e ~ alee, vi^{j}itor ~ viitor, zo^{w}ologie ~ zoologie, no^{w}utate ~ noutate].

Эти два ряда находятся в свободной вариации.

2.2. Гласные [ă, î] после [##] представлены или [^{i}ă, ^{i}î] или же [ă, î]: [^{j}ăla ~ ăla, ^{i}în ~ în].

Эти два ряда находятся в свободной вариации.

2.3. После гласных [e, i, o, u] необходимо различать следующие положения:

а) после [e, i] [a] проявляется или как [^{e}a, ^{j}ă], или же как [a]: [re^{e}al ~ real, arpeči^{j}ă ~ arpečia];

б) после [o] [a] проявляется всегда как [^{o}å, ^{w}å]: [ado^{o}å ~ ado^{w}å];

в) после [u] [a] проявляется или как [^{w}ă], или же как [a]: [lu^{w}ă ~ lua];

г) после [v] [ă] всегда проявляется как [^{w}ă̯]; [do^{w}ă̯];

д) после [u] гласные [ă, î] проявляются или как [^{w}ă̯, ^{w}î] или же как [ă, î]: [lu^{w}ă ~ luă, lu^{w}înd ~ luînd].

Положения **а, в, д** являются случаями свободной вариации, **б** и **г** — дополнительными дистрибуциями.

2.4. После палатального звука ряд [a, o, u] проявляется как [å, ȯ, ů].

2.5. Аллофоны, описанные в пунктах 2—2.3, с точки зрения фонетической, являются гласными, следующими за глайдами, тембр этих глайдов определен или самой гласной, перед которой он стоит, или предшествующей гласной.

Глайды [^{j}, ^{s}] определяют палатальный тембр гласной [a]: [^{e}å, ^{j}ă]; [^{j}] определяет закрытый тембр гласной [e]: [^{j}ẹ], глайды [^{o}, ^{w}] определяют задний тембр гласной [a]: [^{o}å, ^{w}å]; и закрытый тембр гласной [ă]: [^{o}ă̯, ^{w}ă̯]; подобная закрытость производится для [a] звуком [^{i}]: [^{i}ă̯].

3. Что касается редукции вариантов, самые сложные вопросы заключаются в установлении фонологического статуса гласных, предшествуемых глайдами, и фонологического статуса «полугласных».

3.1. Относительно глайдов следует избрать одно из двух возможных фонологических описаний:

3.1.1. Идентифицировать глайды [^{e}, ^{o}, ^{j}, ^{w}] с полугласными [e̯, o̯, j̯, w̯] и рассмотреть группы, составленные гласной, предшествуемой одним из глайдов, как дифонематические; в этом случае будем иметь [jelefant, woperă, wuniversitate, aleje, nowă] и т. д. Неудобство этой интерпретации состоит в том, что невозможно установить тот же фонематический статус для [^{j}] и поэтому нельзя прийти к «pattern congruity».

3.1.2. Рассмотреть те же группы как монофонематические. С помощью этой интерпретации можно прибегнуть к факторам свободной вариации между «глайд+гласная» и «гласная» (без глайда), а также к факторам дополнительной дистрибуции между [a] и [^{o}â, ^{w}ă], между [ă] и [^{w}â]. Избрав это решение, будем иметь [elefant, operă, universitate, alee, noă, real].

Однако здесь имеется трудность, а именно тот факт, что идентичные фонические сегменты получают два различных фонематических статуса.

Если рассматривать пары [mjere — mere], [jăr — ar], [se̯ără — sară], следует придать последовательностям [je, jă, e̯ă] дифонематический статус.

Но те же последовательности получают монофонематический статус при решении, предложенном в п. 3.1.2.

Трудность здесь лишь кажущаяся. Можно рассматривать это положение, как вид нейтрализации на уровне фонематических последовательностей; оппозиция [je — e], например, после [##] нейтрализуется.

3.2. Для полугласных также можно избрать одно из двух возможных фонологических описаний.

3.2.1. Полугласные [e̯, o̯, j̯, w̯] могут рассматриваться, как элементы, отличающиеся от гласных [e, o, j, w], если не рассматривать факторов деления на слоги. То же касается [i̯] в [lupi̯]. Именно так эти факторы рассматривались в большинстве фонологических описаний румынского языка (см. особенно Андрей Аврам).

3.2.2. Я предложил другую альтернативу, которая могла бы учитывать факты деления на слоги. И действительно, оппозиция между [lew] и [le^{w}u] может быть сведена к оппозиции между односложным словом и двухсложным словом. Иными словами можно сказать, что противостоят друг другу не [w] и [u], а **силлабическое деление** (coupe syllabique) между [e] и следующим сегментом и **отсутствие силлабического деления** между [e] и следующим сегментом. Так «оппозиция между единицами сводится к оппозиции между окружением», по выражению З. С. Харриса. Можно рассматривать, значит, [u] и [w] как идентичные, если их считать находящимися в различных окружениях. Различие в окружении состоит как раз в силлабическом делении; в слове [lew] — отсутствие силлабического деления, которое приводит к асиллабическому проявлению звука [u] после гласной. Если между [e] и [u] находилась бы, например, согласная, [w] проявилось бы обязательно как [u]. Именно этот консонантный нуль я предложил обозначить знаком [—] и назвать его «силлабической юнкцией» (англ. juncture).

Таким образом, оппозиция между [lew] и [le^{w}u] является оппозицией между **наличием** и **отсутствием** этого консонантного нуля. Консонантный нуль никогда нельзя смешивать с зиянием. Зияние имеет нечто положительное, так как гласные в зиянии ведут себя как гласные, разделенные согласной (именно поэтому гласные в зиянии развивают согласную, которая заполняет зияние), тогда как юнкция имеет чисто отрицательную природу.

Такая интерпретация должна быть придана всем оппозициям между гласными и полугласными. Каждая полугласная должна быть обозначена

как соответствующая гласная, предшествуемая или предшествующая силлабической юнкции.

Следовательно, можно рассматривать рум. [ę, o̧, j, w] в дополнительной дистрибуции с [e, o, i, u]. Возможность идентификации звуков [ę] с [e], [o̧] с [o], [j] с [i], [w] с [u] была показана в работах фонетической лаборатории Института языкознания в Бухаресте, руководимых А. Росетти.

2.2.3. *Оппозиция между* [i] и [i̯] *также может быть сведена к оппозиции между* [i], *отделенной силлабическим делением от предшествующей гласной, и* [i], *которая от нее не отделена; то есть* [lupi̯ — lupi] *является не чем иным, как оппозицией* [lup — i — lupi]. Знак [—] имеет здесь то же значение консонантного нуля: если вместо [—] находилось бы, например, [l], r], [i] проявлялась бы обязательно, как [i]. Таким образом, можно рассматривать также [i̯] и [i] в дополнительной дистрибуции.

Возможность фонетической идентификации [i̯] с [i] была убедительно показана в экспериментальных работах В. Шутеу: [i̯] — не что иное, как [i] краткое.

2.3. Гласные [à, ò, ù] находятся в дополнительной дистрибуции с [a, o, u]. Именно поэтому необходимо идентифицировать оба ряда.

2.4. Перечень гласных фонем румынского языка таков:

[a] с аллофонами [à, ᵍâ, ʲà, ᵍâ, ʷă];
[e] » » [ʲę, ę];
[i] » » [ʲi, j];
[o] » » [ʷo̧, o̧, ò];
[u] » » [ʷu, w, ū];
[ă] » » [ʲă, ʷą̆, ᵍă];
[î] » » [ʲî, ʷî]

3. *Фонологическое определение этих гласных может быть дано при помощи факторов нейтрализации и морфонологических чередований.*

3.1. Я считаю нейтрализацией всякую дефективную дистрибуцию, разделяющую фонемы на гомогенные ряды с точки зрения фонетической. Единицы, находящиеся в положении нейтрализации, являются экстенсивными (неотмеченными) единицами; другие — интенсивными (отмеченными).

3.1.1. После [o] может появляться только [a], а не какая-либо другая гласная; [a] является гласной без локализации, а другие являются гласными с определенной локализацией. Следовательно, имеются локализованные [e, i, ă, î, o, u], нелокализованная [a].

3.1.2. Гласные [ă, î] не могут следовать за силлабической юнкцией, тогда как другие гласные могут следовать за ней: [ku-i], [mi-ere], [i-ubi], [mi-orlăi]. Гласные [ă, î] являются центральными, другие — нецентральными. Центральные: [ă, î] ~ нецентральные: [e, i, o, u].

3.1.3. Перед [ă, î] приемлемо наличие фонем [o̯, u] и исключено наличие фонем [i, e], то есть в румынском языке возможны только последовательности [oă, oi, uă, ui]; ср. [doă, ploind, luă, luind] и исключены последовательности [iă, iî, eă, eî].

Ряд [i, e] является рядом передних гласных, ряд [o, u] является рядом непередних гласных. Следовательно, нецентральные передние фонемы: [e, i] ~ нецентральные непередние фонемы [o, u].

3.2. В случае чередований (отсутствие оппозиции в данном грамматическом контексте) можно считать, что чередующийся ряд, возможный в данном контексте так же, как ряд, не входящий в чередование, представляет экстенсивный член; ряд, в котором это исключено, представляет интенсивный член.

3.2.1. Гласными чередованиями в румынском языке являются: a ~ e (k'amă ~ kem); e-a ~ e (le-agă — leg); a ~ ă (mare ~ mări);o-a ~ o (sko-ală ~ skol-i); e ~ i (venim ~ vin); e ~ ă (mere ~ măr);i ~ î (vinde ~ ~ vînd); o ~ u (mor ~ murim).

3.2.2. Следует отметить, что чередующийся ряд фонемы [a] состоит из ряда гласных средней открытости и не включает закрытых гласных. Следовательно, можно сказать, что [i, î, u] так же, как [a] могут появляться в грамматических контекстах, из которых ряд [e, ă, o] исключен. То есть [e, ă, o] представляют интенсивный ряд, тогда как [i, î, u] — экстенсивный ряд. Следовательно, можно сказать: открытые [e, ă, o] ~ неоткрытые: [i, î, u].

3.3. Установленная корреляция имеет два измерения: *локализация и открытость*.

3.3.1. В первом измерении имеются следующие корреляции:

а) локализованные: [e, i, ă, î, o, u] ~ нелокализованные: [a];

б) центральные [ă, î] ~ нецентральные: [e, i, o, u];

с) нецентральные передние: [e, i] ~ нецентральные непередние [o, u].

3.3.2. Во втором измерении имеется лишь одна корреляция: открытые [e, ă, o] ~ неоткрытые: [i, î, u].

3.3.3. [a] является членом-абсолютным нулем, поскольку она противостоит всем гласным как экстенсивный член.

В первом измерении следует различать [e, i], которые являются нулевыми членами только первой корреляции и члены [o, u], представляющие нулевые члены обеих корреляций.

Следовательно, в первом измерении (локализация) также есть ряд абсолютного нуля.

Именно этот ряд следует отличать от ряда относительного нуля первой корреляции. Чтобы выразить это различие, обозначим знаком — то, что я назвал относительным нулем и знаком ÷ то, что я назвал абсолютным нулем.

Во втором измерении есть только одна корреляция: [e, ă, o] — открытые ~ [i, î, u] — неоткрытые. Следовательно, можно было бы обозначить интенсивный ряд знаком ÷, а ряд [i, î, u] знаком ø. Но в этом случае нельзя было бы различать [i, î, u], с одной стороны, и [a] (абсолютный нуль по отношению ко всем гласным), с другой стороны. Для сохранения этого различия следует придать ряду [i, î, u] значение ÷.

3.3.4. Теперь мы можем выразить вокализм румынского языка при помощи следующей схемы, в которой на горизонтальной оси расположены места локализации, а на вертикальной оси степень открытости. На каждой оси можно обозначить знаками +, ø, ÷ интенсивные члены, нейтральные и экстенсивные по отношению к данному измерению:

		Центральный		Нецентральный
		+	ø	÷
Открытый	+	ă	o	e
	ø		a	
Неоткрытый	÷	î	u	i

Summary

Upon applying the commutation test the following list is obtained: [*a*, *o*, *u*, *ă*, *î*, *e*, *i*, *ę*, *ǫ*, *j*, *w*, ^{w}u, ^{j}i, ^{i}i].

The subsequent series of alophones are described by means of transformation rules: [$^{j}ę$, ^{j}i, $^{w}ǫ$, ^{w}u] are in free variation with [*e*, *i*, *o*, *u*] after pause or vowel; [$^{i}ă$, $^{i}î$] are in free variation with [*ă*, *î*] after pause; [$^{ę}ȧ$, $^{j}ȧ$], are in free variation with [*ȧ*] after [*e*, *i*]; [$^{ǫ}å$, $^{w}å$], after [*o*, *u*]; [$^{w}ă$, $^{w}î$], after [*o*, *u*]; [*ȧ*, *ȯ*, *u̇*] after palatal sounds.

The alophones [$^{j}ę$, ^{j}i, $^{w}ǫ$, ^{w}u, $^{j}ȧ$, $^{ǫ}å$, $^{w}å$, $^{w}ă$, $^{w}î$] are interpreted as monophonematic segments. The non-syllabic segments [*ę*, *ǫ*, *j*, *w*] are interpreted as alophones of the vowels [*e*, *o*, *i*, *u*], respectively, admitting the existence of a syllabic juncture [—].

The inventory of the vocalic phonemes is:

[*a*]	with	the	alophones	[*ȧ*, $^{ę}ȧ$, $^{j}ȧ$, $^{ǫ}å$, $^{w}å$]
[*e*]	»	»	»	[$^{j}ę$, *ę*]
[*i*]	»	»	»	[^{j}i, *j*]
[*o*]	»	»	»	[*ȯ*, $^{w}ǫ$, *ǫ*]
[*u*]	»	»	»	[*u̇*, ^{w}u, *w*]
[*ă*]	»	»	»	[$^{i}ă$, $^{w}ă$, $^{ǫ}ă$]
[*î*]	»	»	»	[$^{i}î$, $^{w}î$]

Upon examining the data of vocalic alternation and neutralization, the following system is attained:

		Central		Non-central
		+	ø	÷
open	+	ă	o	e
	ø		a	
close	÷	î	u	i

Р. Г. Пиотровский

СОСУЩЕСТВУЮЩИЕ ФОНЕТИЧЕСКИЕ СИСТЕМЫ И СТИЛИСТИЧЕСКИЕ КОРРЕЛЯЦИИ В МОЛДАВСКОМ ЯЗЫКЕ

Настоящая статья опирается на двухступенчатую теорию фонологии, разработанную С. К. Шаумяном. Согласно этой теории следует различать ступень наблюдения (фонетическая ступень) и ступень конструктов (фонематическая ступень)[1]. В качестве единицы фонетической ступени выступает фонемоид. Структурно организованное множество фонемоидов называется фонетической системой.

Звуковой строй молдавского языка охватывает несколько сосуществующих фонетических систем, ориентированных на определенные стили речи.

Высокие стили

Вокализм высоких стилей[2] молдавского языка стремится к девятичленной системе, характерной и для румынского литературного языка.

и	ы	у (ӳ)[3]
е	э	о (ӧ)[3]
‘ä	а	"а ("ä)[3]

1. В отличие от остальных стилей, [и] и [е] выступают здесь как самостоятельные фонемоиды, соответственно противопоставленные фонемоидам [ы] и [э]. Ср. корреляции: *ин : ын*, (*йеу*) *йнтру : ынтру; мер : мэр, пер : пэр*[4].

2. В отличие от средних и низких стилей, в современном «высоком» молдавском произношении характерно «суперкорректное» введение в группу «смягченная согласная+гласная» полугласного элемента, в результате чего образуется отчетливо слышимый дифтонг. Ср. произношение *г'ӧл, л'ӓк, п'ӓтрэ*.

---

[1] С. К. Шаумян, Проблемы теоретической фонологии, М., 1962, стр. 38—46.

[2] Высокие стили речи охватывают торжественно-поэтический и научный обиход. Речь идет об устной форме существования этих стилей — т. е. о чтении стихов вслух — чаще всего на сцене, об устном выступлении с научным докладом или лекцией.

[3] В скобках указывается палатализованный вариант гласной, употребляющийся после смягченной согласной (ср. кэ|дуц| : у|л'ӳц|). Эти палатализованные варианты встречаются очень редко (это относится и к неударяемым позициям). Что касается минимальных пар, в которых обнаруживается оппозиция этих вариантов в связи с противопоставлением твердых и мягких согласных, то их число не превышает пятишести случаев.

[4] В остальных стилях эти корреляции не выдерживаются то ли в результате фонетических изменений (*ин*>*йин*, *йнтру*>*йнтру*), то ли в результате неупотребительности соответствующих слов (*мер*, *пер*).

3. В высоких стилях смягчение согласных в середине слова не имеет фонологической функции и является избыточным элементом (не случайно, что палатализация гласных перед [и], [е] и [ʲа] здесь слабее, чем в средних и низких стилях).

Фонологическое противопоставление твердых и смягченных согласных обнаруживается лишь на конце слова [5]. Оно охватывает следующие согласные:

п б м ф в т н з ц ш ж л р к г х
п' б' м' ф' в' т' н' з' ц' ш' ж' л' р' к' г' х'.

Не имеют смягченных фонемоидных пар согласные [д, с], а твердых — согласные [ч', дж'] и [й].

Средние стили [6]

Для вокализма средних стилей характерны две важные тенденции:

а) устранение фонологической корреляции [и : ы, е : э] (см. выше) [7].;

б) падение полугласного элемента в дифтонгах [ʲä, ʷä, ʲö, ʷў], в результате чего последние превращаются в простые гласные, а смягченность согласных из избыточного элемента превращается в дифференциальный признак.

В результате указанных тенденций система гласных в средних молдавских стилях образует шестичленный четырехугольник, реализующийся в двух комбинаторных вариантах:

---

[5] С акустической точки зрения здесь следует говорить о комбинации смягченного и твердого согласного с особого вида «призвуками» или «окрасками» (ср. E. P e t r o v i c i. Influenţa slavă asupra sistemului fonemelor limbii române. Bucureşti, 1956, стр. 20). Как показали спектрографические исследования молдавской речи, эти призвуки имеют консонантный характер. Этому вопросу посвящена специальная статья.

[6] Средние стили включают литературно-повествовательный и литературно-разговорный стили.

[7] Начальные [и] и [е] в средних стилях обычно йотируются даже в иностранных словах: *йекаám'ен*, *йизотóп*, *а йед'итá* и т. п.

[8] Ср. рифмы Эминеску: *аржинт — мормынт* («Ежипетул»), *фынтыне — мине* («Сара пе дял»), *омят — млэдиет*. А. Аврам (см.: Consideraţii fonologice asupra rimelor româneşti. «Fonetică şi Dialectologie». Bucureşti, 1958, стр. 9—16) считает, что рифмы этого типа не должны считаться унифонемными, но являются простыми созвучиями. С этим положением вряд ли можно согласиться, поскольку для рифм Эминеску характерен в подавляющем большинстве случаев фонемный (по нашей терминологии фонемоидный) параллелизм.

[9] Говоря о самостоятельном дифтонге [ᵘа] [ᵒа], а не о лабиальной корреляции мы исходим из следующих соображений:

1. Лабиализация согласных типа *майᵛ*, *соцᵛ* в молдавской речи встречается редко и имеет чисто диалектальный характер (район Измаил-Рени).

2. При устном (не школьном) усвоении молдавского языка не молдаванами (русскими, украинцами, болгарами) (дифтонг [ᵒа] произносится обычно как [о] (ср. *кóдэ* вм. *кᵒáдэ*, *тоте* вм. *тᵒáтэ* и т. п.), что говорит о том, что [о] воспринимается не как лабиализованная окраска согласного, но как гласный элемент, более сильный, чем вторая часть дифтонга — [а].

3. Монофтонгизация дифтонга [ᵒа] в [о] отмечается и в некоторых молдавских говорах (не только на Буковине и в Закарпатье, но и в Бессарабии). Ср. диалектные ошибки учащихся средних школ: *се сколэ* вм. *се сковлэ*, *форте* вм. *фоарте* (Братушанский район); *грозникэ* вм. *гроазникэ* (Кишиневский р-н); *поте* вм. *поате*, *улчоре* вм. *улчоаре* (Дубоссары). Одновременно в современном разговорном и диалектном произношении в некоторых употребительных словах обнаруживается тенденция к монофтонгизации дифтонга [ᵒа] не в [о], но в простое [а] (без лабиализации предшествующего согласного): *кулᵒáре* > *колáр'е*; *пᵒáте* > *пáте*; *тᵒáт'е* > *тáте*, *тэте*; *фᵒáрте* > *фáрте*. Ср.: A. G r a u r. Les diphtongues en roumain. BL, III, 1935, стр. 45, I. I o r d a n. Note Lingv. BPh, VI, 1939, стр. 139; E. S e i d e l. Linguistische Beobachtungen in der Ukraine. BL, XI, 1943, стр. 108.

I. Гласная стоит после твердой согласной или в начале слова:

ы[8]	у
э	о
а	ъа[9]

II. Гласная стоит после мягкой согласной:

и	ӳ (орфогр. *ю*)
е	ӧ (орфогр. *ьо, йо*)
ӓ (орфогр. *я*)	ъа (орфогр. *ьоа*)

Фонетическая система гласных средних стилей выявляется при ударенном употреблении гласных, а также при употреблении их в односложных словах. Однако даже в ударенном положении, не говоря уже о неударенных позициях, здесь отмечается тенденция к дальнейшему упрощению этой системы, проявляющаяся в смешении гласных фонемоидов: *рыд* : *рад*, но *рынта́ш* = *ранта́ш; выр* : *вэр*, но *кэлкы́й* = *кэлкэ́й, тылха́р* = *тэлха́р; н'и* : *н'е*, но *тр'им'и́те* = *тр'им'е́те; б'ӓт* : *б'ет*, но *б'ӓ* = *б'е*[10].

Для консонантизма средних стилей характерно два противоположных процесса. С одной стороны, в результате монофтонгизации дифтонгов смягченность предшествующих согласных становится различительным признаком в парах типа *г''ӧл* : *гол, л''ӓк* : *лак, п'ӓтр*(*э*)[11] : *пӓтр*(*у*).

С другой стороны, уже в средних молдавских стилях прослеживается отчетливая тенденция к устранению корреляции «твердые: смягченные согласные» как в середине, так и на конце слова. Это устранение осуществляется очень неравномерно и в лингвостилистическом и в лингвогеографическом аспектах [12]. В средних стилях этот процесс проявляется:

а) В обобщении смягченного [р'] в некоторых положениях на конце слова: *мора́р'* : *мора́р'* [13]; *мэга́р'* : *мэга́р', стежа́р'* : *стежа́р'* вм. *мора́р* : *мора́р', мэга́р* : *мэга́р', стежа́р* : *стежа́р'*.

б) В обобщении твердого члена коррелятивной пары. Это касается в первую очередь сибилянтов [с, з] — *(сэ)са́рэ* : *са́рэ, лукр'е́з* : *лукр'е́з*, вм. *(сэ)са́рэ* : *с'ӓрэ*[14], *лукре́з* : *лукр'е́з;* шипящих [ш, ж] — *кукӧш* : *кукӧш, г'ӳж* : *г'ӳж* вм. *кукӧш* : *кукӧш', г'ӳж* : *г'ӳж'* и аффрикаты [ц] — *соц* : *соц* вм. *соц* : *соц'* [15].

Процесс отвердения охватывает не только сибилянты, шипящие и [ц], но и другие согласные. Ср. высокие стили *кр'ъ́нгэ* — средние стили *крангэ*.

Вместе с тем в средних молдавских стилях отмечается тенденция к ослаблению смычного элемента в аффрикатах [дж'] и [ч'] в результате чего [дж'] превращается в мягкий, продвинутый вперед [ж''], а [ч']

[10] Ср. орфографические колебания у Крянги: *кэлкый — кэлкэй, кэпэтый — кэпэтэй, тримите — тримете — тримятэ — тримитэ, бе — бя, ave — авя, спэриета — спэриатэ* — см. Gr. Scorpan. Note Lingv. Limba lui Creanga, BPh, I, 1934, стр. 104—109; или у Эминеску: *рэскоалэ — рескоалэ* — см.: Gr. Scorpan. Note Lingv. BPh, III, 1936, стр. 121—122 прим., *спэриет — сперият* («Диана») — см. там же.

[11] С исторической точки зрения, произношение *п'а́трэ* является сравнительно новым гиперурбанизмом: *п'ъ́трэ* > *к''а́трэ* > *п'ӓтрэ* см. ниже.

[12] Ср.: Р. Г. Пиотровский. Некоторые теоретические проблемы Молдавского лингвистического атласа, «Volumul omagial I. Iordan». Bucureşti, 1958, стр. 679—680.

[13] Исторически [р'] является здесь более старой формой. Ср. молд. [-ар'] < лат. arius.

[14] Ср. орфографические колебания *сарэ — сярэ* у Крянги (BPh, II, стр. 104) и у Эминеску: *Сара пе дял — кэч есте сяра'н* асфинцит («Лучафэрул»).

[15] «Ортография лимбий молдовенешть». Кишинэу, 1957, стр. 17. Ср. неразличение твердости и смягченности указанных согласных в рифмах Эминеску типа: *заплаз — авь, чиреш — ешь, кожь — рэбож, колц — болць*, — см. Gr. Scorpan. Note Lingv. BPh, III, 1936, стр. 120—121.

приближается иногда к мягкому продвинутому [ш'']. Впрочем, в последние годы в средних молдавских стилях отмечается тенденция к восстановлению [ч'] и отчасти [дж']. О суперкорректных [дж'] и [ч'] см. ниже.

Низкие стили

(просторечье, смыкающееся с говорами Правобережных районов МССР)

В ударенном положении система гласных та же, что и в средних стилях.

В неударенных положениях система гласных еще более упрощается.

1. «Мягкие» комбинаторные варианты [ӓ], [ᵒа] в неударенных положениях обычно не употребляются.

2. В абсолютном исходе слова отмечаются следующие явления:

а) В отличие от средних и высоких стилей неударенное [е] после мягких согласных произносится в просторечье как [и].

Ср. *н''и-й спу́н'и* (*м''и-й спу́не*), *фра́т'и*, *пы́н'и* и т. д., ср. ошибки учащихся типа *мынили* (вм. *мыниле*) (Дубоссары).

б) [е] после твердых согласных и [э] произносятся у большинства носителей просторечья как слабое [ы]: *ка́сы* : *ка́сы*, *ф'ет'и́цы* : *ф'ет'и́цы* вм. *ка́сэ* : *ка́се*, *ф'ет'и́цэ* : *ф'ет'и́це*[16] и т. д., а иногда и вовсе исчезает: *кас*, *мэтӯш* вм. *ка́се*, *мэтӯшэ*. Ср. орфографические ошибки учащихся типа *одат* (вм. *одатэ*) (Дубоссары); *лукраре скрис* (вм. *скрисэ*) (Вертюжены). В результате этих изменений система гласных в конечной неударенной позиции выглядит так:

ы/и	у/ӳ[17]
—	о/ӧ
а	

3. В середине слова в неударенных положениях отмечаются тенденции к постоянному смешению произношения следующих гласных: [е] и [э], [е] и [и], [а] и [э], [э] и [о], реже [и] — [у] и [о] — [у]. Ср. произношение *зе́ч'и* — *зэ́ч'и*, *ф'е́тел'е* — *фе́т'ил'и*, *ф'ем'е́и* — *ф'им'е́и* — *фум'е́и*, *ше́пт'и* — *ша́пт'и* — *ша́пт'и*, *пэпшо́й* — *попшо́й*, *батры́н* — *бэтры́н*, *нэро́к* — *норо́к*, *поро́нкэ* — *поры́нкэ* и т. п.[18]

Особенно слабой оказывается позиция предударного гласного, который обнаруживает тенденцию к выпадению. Ср. произношения: *попшо́й*, *Кишнэ́у* и ошибки типа *молдовняскэ* вм. *молдовеняскэ* (Вертюжены).

В результате этих смешений вся система неударных гласных в середине слова дает следующую схему:

ы(э)/и(е) у/ӳ
∧(а, е, э, о)

Просторечный консонантизм характеризуется дальнейшим устранением противопоставления твердых и мягких согласных. Кроме отвердения сибилянтов, шипящих и аффрикат, смягчения конечного [р] после [а, е, о] — процессов, известных в средних стилях, для просторечья характерна палатализация смягченных губных [19].

---

[16] Ср. рифмы Эминеску типа: *фэлоасе* — *луксоасе* — *апасэ*, *браце* — *фаце*, *разэ* — *инчетязэ*, — см.: Gr. Scorpan. Op. cit., стр. 120.

[17] [и, ӳ] используются после смягченной, [ы, у] после твердой согласной.

[18] Ср. у Крянги: *пэпушой* — *попушой*, *Некулай* — *Николай*, *ынфэцишя* — *ынфэцоша* (BPh, I, стр. 104—105). Просторечные варианты указанных выше типов встречаются и в речи героев Караджале, где они служат средством передачи молдавского колорита. См.: I. Iordan. Limba «eroilor» lui I. L. Caragiale. Bucureşti, 1957, стр. 15, 16 и 23.

[19] Отдельные палатализованные формы квалифицируются уже как литературные, а иногда фигурируют и в словарях — ср. *кирон* (<пирон), *кипэруш* (<*пипэруш*).

В результате этого явления противопоставление губных согласных по твердости и смягченности сменяется корреляциями иных типов — например переднеязычный: среднеязычный *пом* : *пон''* вм. *пом* : *пом'*; *вэтáв* : *вэтáж''* вм. *вэтáв* : *вэтáв'* или переднеязычный: заднепалатальный *боб* : *бог''* вм. *боб* : *боб'*, *бэт* : *г''ет* вм. *бэт* : *б'ет*, *луп* : *лук''* вм. *луп* : *луп'*[20].

Кроме того, для просторечья характерна:

а) полная утрата смычного элемента в аффрикате [ч'], в результате чего последняя превращается в мягкий продвинутый [ш'']: *ч'ер* > *ш''ер'*;

б) сохранение в словах определенных этимологических классов [дз] вместо [з]: *лукрéз* > *лукр'éдз*.

Итак, звуковой строй современного молдавского языка охватывает три сосуществующих фонетических системы, каждая из которых ориентирована на определенный стиль речи. Соотнесенность этих фонетических систем осуществляется через стилистические корреляции звуковых единиц (фонемоидов и их вариантов), находящихся в однозначных с синтагматической точки зрения позициях — ср. синонимический ряд: выс. стиль (*ту*) *лукрéз'*: средн. стиль (*ту*) *лукр'éз*: простор. (*ту*) *лукр'éдз*. Каждый из элементов этих стилистических корреляций мы будем обозначать термином «стилистема» (само собой разумеется, что стилистемой может быть и грамматическая форма, и слово).

Таким образом, каждая фонетическая единица в сосуществующих системах оказывается коррелированной по двум осям — фонологической и стилистической.

Но как соотносятся и взаимодействуют между собой эти «оси»?

Для этого рассмотрим стилистические противопоставления фонемоидов [з'] и [з(дз)] в паре *еу лукрез* : *ту лукрезь*:

Высокие стили	(*еу*) *лукрéз*	(*ту*) *лукрéз'*
Средние стили	(*еу*) *лук*[*р'*]*éз*	(*ту*) *лук*[*р'*]*éз*
Просторечие	(*еу*) *лукр'éдз*	(*ту*) *лукр'éдз*

Упрощая эту схему, получаем:

стилистика \ фонология	з	:	з'
	‖		..
	з	=	з
	..		..
	дз	=	дз

где

: — фонологическое противопоставление,

= — фонологическое равенство

.. — стилистическое противопоставление

‖ — стилистическое равенство

---

[20] Возникающие в результате палатализации губных среднеязычные и заднепалатальные рефлексы следует рассматривать не как позиционные варианты мягких губных, но как отдельные фонемоиды. Об этом свидетельствует:

а) возникновение новых корреляций типа (*йеу*) *в'ин*: *γ''ин*, *ж''ин* вм. (*йеу*) *в'ин*: *в'ин*; (*ел сэ*) *в'ие*: *γ''ие*, *ж''ие* вм. (*йел сэ*) *в'ие* : *в'ие* (заметим, что отдельные говоры дают палатализацию [в'] в формах глагола *а вени* — ср. (*йеу*) *γ''иу* (Рени);

б) ложная суперкорректная этимологизация палатализованных форм типа *ш''ер* < *ч'ер* вм. *ф'ер*; *ж''ин* > *дж'ин* вм. *в'ин*, свидетельствующая о разрыве фонологической связи между смягченным губным и его среднеязычным рефлексом.

Схемы стилистических корреляций некоторых молдавских фонемоидов и их вариантов

Фонемоиды и их варианты	Стили		
	высокие	средние	низкие
Губной в сочетании с дифтонгами [^{u}ä, ^{u}е] Мягкий [з']	п'^{u}а́трэ, б'^{u}ет: з'и, з'^{e}а́мэ, (ту) лукре́з'	п'а́трэ, б'ет зы, за́мэ, (ту) лукр'е́з	к''а́тры, г''ет дзы, дза́мы, (ту) лукр'е́дз
Начальные [и, е] Дифтонги [^{u}ö, ^{e}ä] Мягкие сибилянты, шипящие и аффриката [ц] [дж']	институ́т, екза́мен г'^{u}ӧл, л'^{e}ӓк с'^{e}а́рэ, кукóш', соц' дж'ер		йипст'иту́т, йекза́м'ен (йикза́м'ин) г''ӧл, л''ӓк са́рэ, кукóш, соц ж''ер
Конечный неударенный [е] [ч'] [з]		би́не, пе, пы́не (б'и́не, пы́н'е) ч'и́не, (ч'и́н'е) (йеу) лукре́з: (йеу) лукр'е́з	г''ини, п'и, пы́н'и ш''ин'и (еу) лукр'е́дз

Ср. также другие типы фонетико-стилистических корреляций в молдавском языке (см. схему).

Анализ приведенных корреляций позволяет сделать следующие заключения:

1. Стилистическая корреляция может осуществляться как путем фонетических противопоставлений *(б')ин'е : г''ин'и, лукрé(з') : лукр'é(з)*, так и посредством противопоставлений их вариантов *бúне : б'ин'е, лукрéз : лукр'éдз*. В связи с этим в качестве стилистем могут выступать как фонемонды, так и их варианты.

2. Принципы фонологической классификации не совпадают с принципами стилистического противопоставления[21]. Фонемонды (и их варианты) одного и того же ряда могут давать различное стилистическое распределение. Ср. корреляции: высокий стиль [дж']: средний стиль и просторечие [ж''] и высокий и средний стили [ч']: просторечие [ш'']. Наоборот, разные фонемы (и их варианты и комбинации) дают одинаковое стилистическое распределение: [п'ъä. . .: п'ä. . .: к''ä. . .] = [з' : з : дз].

* * *

Одной из характерных черт современного молдавского разговорного стиля является частое использование псевдокорректных форм типа *ч'и* вм. *ч'е, б'ин'и* вм. *бúне, зэч'и* вм. *зéч'е, в'ицэл* вм. *в'ицéл* и т. п. Отличительной чертой этих форм является восстановление литературного произношения согласной при сохранении просторечной формы гласного (ср. соответствующие просторечные *ш''и, г''ин'и, дзэши, ж''ицэл*). Современные литературные формы с мягкими губными согласными также являются результатом древней псевдокорректной реституции: *п'ъ'ътрэ > к''ътрэ > п'ътрэ, ф'''ер > х''ер > ф'ер* и т. п.[22]

Вообще для носителей молдавской речи характерно ясное осознание стилистических корреляций согласных и «небрежное» невнимание к менам гласных — особенно в неударенных позициях. Ср. приведенные выше примеры типа *ф'ем'се — ф'им'еи — фум'еи*. Об этом говорят также следующие факты:

1. Подавляющее большинство орфографических ошибок у молдавских школьников падает на употребление гласных (ср. написания *вистит* вм. *вестит*, *шидя* вм. *шедя*, *с'а мынтуит* вм. *с'а мынтуит*, *вазут* вм. *вэзут* и т. д. и т. п.), согласные дают меньшее количество ошибок (в основном это смешение твердых и мягких типа *пе немц* вм. *пе немць* и *ун нямць* вм. *ун нямц*, а также оглушение согласных типа *разветкэ* вм. *разведкэ*; ошибки на палатализацию губных практически отсутствуют)[23].

2. У писателей-классиков отмечаются многочисленные случаи орфографической вариантности при передаче гласных — в том числе и ударенных (ср. примеры из произведений Эминеску и Крянги на стр. 93—95). Просторечные отклонения от литературной нормы в области консонантизма строго ограничены и имеют, как правило, стилевую нагрузку — см. у Садовяну в прямой речи крестьянина: *acu-i de cherit* (вм. *perit*), *cucoane Cilipe* (вм. *Filipe*) (*Venea o moara pe Siret*, Bibl. p. toți, Bucureşti, 1956, стр. 10).

---

[21] Ср.: Р. Г. Пиотровский. Очерки по стилистике французского языка, изд. 2-е. Л., 1960, стр. 9.

[22] W. Meyer-Lübke. Zentripetale Krafte im Sprachleben. «Hauptfragen der Romanistik. Festschrift Ph. A. Becker», 1922, стр. 138; D. Macrea. Palatalizarea labialelor în limba romînă.

Ср. в этом плане гиперурбанизмы *nifilist* (<*nihilist*), *arfiva* (<*arhiva*), используемые у Караджале. BPh, I, 1933, стр. 156.

[23] Ср.: G. Istrati. Graiul popular seris. BPh, VI, 1937, стр. 139.

3. Орфографическая и произносительная вариантность в молдавских топонимах также имеет, как правило, вокалический характер — *Болбока — Бульбока*, *Карбуна — Кэрбуна*, *Корнешты — Кырнешты*, *Маринешты — Меринешты* и т. п. Колебания в согласных сравнительно редки — *Бравича* : *браγ''и́ч'а*, *Випрова* : *γ''ипрова*, *Филипень* : *х''ил'ип'ён'*, *ш''ил'ип'ён'*, возможно *Кишинэу* < **п'иш'инэ́у*[24].

Все эти факты свидетельствуют о различной информационной нагруженности молдавских согласных и гласных. Первые образуют в слове несущий информацию костяк [25], вторые получают, как видно, меньшую информационную нагрузку.

В связи с этим стилистические корреляции согласных [б' : г''], [ч' : ш''], [з' : з : дз͡] выступают как интенсивно маркированные; наоборот, вокальные стилистические корреляции [е : и, п : ы, э : ы, э : а : о] оказываются слабо маркированными, а иногда и вообще не маркированными.

Отсюда становится ясным, почему передача жанрово-стилистической (социально-лингвистической [26]) информации в ее различных формах (гиперурбанизм, речевая характеристика героев художественных произведений) осуществляется в молдавском языке чаще всего с помощью согласных.

COEXISTENT PHONEMIC SYSTEM AND STYLISTIC CORRELATIONS IN MOLDAVIAN

Summary

The purpose of this paper is to apply the structural method to a stylistic study. The coexistent phonemic systems form the phonological superstructure of a language. At the same time every phonemic system is connected with a definite style of this language. Thus every phonemic unit is correlated up with two axes — a phonological axis and a stylistic axis.

---

[24] Ср.: Г. Ф. Богач. Заметки по молдавской топонимике и идиоматике. «Изв. МФАН» 1957, 12.

[25] Ср.: G. E. Peterson. Information Bearing Elements of Speech. «Journ. of Acoust. Soc. of Am.», XXIV, 1952, стр. 629—637.

[26] Ср.: P. Ladefoged and D. E. Broabent. Information convoyed by vowels. «Journ. of Acoust. Soc. of Am.», XXIX, 1957, стр. 98.

III. СТРУКТУРНАЯ ГРАММАТИКА

С. Я. Фитиалов

О МОДЕЛИРОВАНИИ СИНТАКСИСА В СТРУКТУРНОЙ ЛИНГВИСТИКЕ

Язык можно описывать с двух точек зрения — грамматической и семантической. Грамматическое описание характеризует язык «изнутри», семантическое — «извне».

Оба вида описания тесно связаны друг с другом. При разработке объективных и точных методов описания языков нужно достаточно ясно и недвусмысленно учитывать эту связь.

В данной работе делается попытка уточнения терминов «семантика» и «грамматика»; на основе этого уточнения изучаются возможные типы грамматического (синтаксического) описания языков аксиоматическими средствами.

1. СЕМАНТИКА

1. 1. В самом общем случае любой язык, естественный или искусственный, можно рассматривать как некоторый код, с помощью которого представляются и передаются сообщения. При этом сообщения кодируются конечными последовательностями (цепочками) знаков из заданного конечного алфавита.

Поскольку код связан с сообщениями, при его полном задании или описании необходимо указывать соответствие между сообщениями и кодовыми цепочками. Это соответствие определяет с е м а н т и к у кода. В простейшем случае семантика кода может быть задана таблицей, сопоставляющей кодовые цепочки и сообщения (если число тех и других конечно).

Как правило, для интересующих нас кодов не любая цепочка в алфавите кода может быть кодовой цепочкой, т. е. соответствовать некоторому сообщению. Могут существовать цепочки, которые с точки зрения семантики кода «бессмысленны» — им не соответствует никакое сообщение. В этом случае говорят, что код избыточен. Естественные языки обладают большой избыточностью.

Избыточные коды можно описывать «изнутри», выделяя из всевозможных цепочек только кодовые цепочки, «осмысленные» при заданной семантике кода.

1. 2. Описание семантики естественных языков как кодов в указанном смысле является чрезвычайно трудной задачей. Здесь сразу возникает проблема определения, что такое «сообщение». Реально, при попытке точного задания семантики языка мы всегда соотносим его с некоторым другим (естественным или искусственным) языком, семантика которого в точном смысле этого слова также предварительно не задана. При соотнесении этих языков друг с другом один из них определяет семантику дру-

гого и наоборот, т. е. мы имеем замкнутый круг. Соотнесение осуществляется следующим образом. Каждый язык описывается самостоятельно «изнутри», т. е. аксиоматически определяются кодовые цепочки языка. После этого сообщениями с точки зрения первого языка считаются кодовые цепочки второго языка и наоборот. Устанавливается соответствие между кодовыми цепочками двух языков, которое и определяет их семантику. Пригодность аксиоматики кодовых цепочек и соответствия между языками должна проверяться практически, с точки зрения поставленной цели.

1. 3. Изложенное понимание семантики языка, основанное на рассмотрении языка как кода и на соответствии между кодовыми цепочками двух языков, может показаться несколько спорным и ограниченным и даже навести на мысль, что произошла подмена термина. Надо сказать, что автору неизвестно никакое достаточно определенное и общепринятое понимание термина «семантика», и поэтому здесь происходит не подмена термина, а его уточнение, пригодное для достаточно большого круга задач: для точного определения отношения перевода между языками и построения систем формального («машинного») перевода, для выработки объективных методов обучения иностранному языку, для задачи информационного накопления, т. е. для задач, в которых нас интересуют соответствия между языками, естественными или искусственными. Для этих задач понимание семантики как «семантики соответствия» или «семантики перевода», по-видимому, может оказаться достаточным.

1. 4. Существуют ситуации, в которых трудно говорить о семантике языка как о соответствии между ним и другим языком. Простейший пример — овладение языком в первые годы жизни человека, не знающего еще ни одного языка, с которым можно было бы соотнести изучаемый язык.

Имея в виду такие ситуации и связанные с ними задачи, целесообразно пользоваться другим пониманием семантики языка, который рассматривается уже не как код, а как система реакций организма на воспринимаемые явления окружающего мира в самом широком смысле этого слова. Такое понимание семантики должно быть тесно связано с проблемой обучения (в кибернетическом плане).

1. 5. Полезно разграничивать эти два различных подхода к пониманию семантики языков. Для структурной лингвистики, по-видимому, типично понимание семантики как «семантики соответствия». Фонологические и грамматические описания языка имеют своей целью построение внутренней аксиоматики допустимых кодовых цепочек; отношения перевода между языками, основанные на такой аксиоматике, описывают семантику языков относительно друг друга.

При понимании семантики как «семантики реакций» мы вынуждены выйти за пределы собственно лингвистики и углубиться в область философии, психологии, физиологии и других наук, руководствуясь общекибернетическими принципами.

Безусловно, при построении внутренней аксиоматики языков и соответствий между ними могут использоваться оба понимания семантики, как и любые другие средства. Нас будут интересовать не способы построения внутренних описаний языков, а сами эти описания.

2. ГРАММАТИКА

2. 1. С точки зрения внутреннего описания язык является множеством конечных цепочек в заданном алфавите букв или звуков, которые интуитивно классифицируются человеком, владеющим языком, как «правильные», или «грамматические», или «осмысленные» цепочки.

Имея в виду для определенности языки с буквенной письменностью обычного типа, можно выделить три типа цепочек, образующих «правильные» последовательности букв языка: с л о в о ф о р м ы (последовательности букв между пробелами или знаками препинания), п р е д л о ж е н и я (последовательности словоформ между точками или прописными буквами) и т е к с т ы (последовательности предложений).

Грамматика языка, или его внутреннее описание, представляет собой совокупность правил, определяющих «правильные» последовательности букв языка и отделяющих их от «неправильных» последовательностей. Такое понимание грамматики было сформулировано Хомским.

2. 2. При построении грамматики языка возникает трудность, кажущаяся непреодолимой. Даже хорошо знающему язык человеку не всегда ясно, что считать правильным, а что неправильным. Приведем несколько примеров. Для русского языка цепочки *вщскпр*, *в в в*, *за человек стол*, *человек за сидит столом* несомненно неправильны. Но цепочки *обсуждались завершенные работы писателями* и *представленные картины на выставку пользовались большим успехом* могут с первого взгляда вызвать некоторое смущение [1]. Если мы и считаем их «неправильными», то далеко не по интуитивным соображениям — ведь они вполне «осмысленны» и «понятны».

2. 3. Бессмысленно строить грамматику для всего языка, тем более что неясен смысл термина «весь язык». Мы можем построить грамматическую аксиоматику так, что приведенные выше два «сомнительных» предложения будут определяться как «правильные», но можем поступить и наоборот. Все зависит от того, для какой цели строится грамматика. Например, грамматика, предназначенная для построения выходных предложений при переводе с одного языка на другой, должна быть достаточно «строгой» для того, чтобы исключить возможность построения «плохих» предложений — небольшая грамматическая или стилистическая «неправильность» может исказить содержание переводимого текста, сделать перевод не вполне понятным или двусмысленным. С другой стороны, если мы строим грамматику, предназначенную для анализа исходного текста при его переводе, то мы должны следить в основном только за тем, чтобы не отнести к «неправильным» предложения, которые по ряду причин реально встречаются в текстах; если же при этом в число «правильных» предложений будут включены заведомо «неправильные» цепочки, то это не помешает использованию нашей грамматики при анализе текста — ведь такие цепочки никогда не встретятся.

2. 4. Безусловно, построение грамматик, описывающих язык наиболее полно и точно как со стороны правильности, так и со стороны неправильности — задача важная и нужная. Однако вряд ли можно будет когда-нибудь остановиться на одной, «самой хорошей» грамматике — достаточно вспомнить о так называемой разговорной речи и об изменениях, происходящих в языке с течением времени. Кстати, последнее явление постоянно будет приводить к устареванию грамматик.

2. 5. Грамматика, отделяющая «правильные» цепочки букв от «неправильных», может быть задана тривиальным способом — списком, если язык конечен. Однако естественные языки вряд ли целесообразно считать конечными (по крайней мере с практической точки зрения). В случае бесконечных языков построение грамматики, которая должна быть конечной и достаточно обозримой, является далеко не тривиальной задачей.

---

[1] Предложения подобного типа, встречающиеся в некоторых статьях и радиопередачах, приводятся в статье Н. Долининой «Маскарад слов» («Известия», 29 ноября 1960 г.).

При предварительном изучении теоретически возможных грамматик естественных языков полезно идти не только по пути построения конкретных грамматик, но и по пути выяснения общего характера грамматик, пригодных для описания языков.

Грамматику можно условно разделить на три части: морфологию, определяющую словоформы языка, синтаксис, описывающий предложения языка в терминах этих словоформ, и описание, определяющее тексты языка в терминах предложений.

Практически (по крайней мере на первых порах) при описании грамматической структуры языка множество словоформ можно считать конечным. В этом случае построение морфологии не представляет принципиальных трудностей — ее можно всегда задать списком.

В первом приближении синтаксис можно строить на основе заданного конечного перечисления словоформ. Содержательный и практически полезный синтаксис должен использовать некоторое точное понятие структуры предложения, на котором и основывается описание «правильных» предложений языка.

Что касается описания «правильных» последовательностей предложений, то эта задача еще не ставилась в целом, хотя элементы «грамматики текстов» и рассматривались (например, соотнесение местоимений и замещаемых ими словоформ и конструкций в разных предложениях).

2. 6. Определение предложений языка можно давать двумя способами. Первый способ заключается в указании правил, с помощью которых можно строить цепочки словоформ; предложения языка определяются как цепочки, которые можно построить по данным правилам. Если цепочку словоформ построить таким образом невозможно, то она не считается предложением языка. Синтаксис такого типа представляет собой и с ч и с л е н и е — совокупность правил, указывающих, что м о ж н о делать при построении предложений языка. Структура предложения определяется способом его построения (выводом) в данном исчислении.

Исчисление называется р а з р е ш и м ы м, если можно построить алгоритм, который для каждой данной цепочки словоформ определяет в конечное число шагов, является ли она предложением языка, описываемого исчислением, или нет. Такой алгоритм восстанавливает структуру цепочки (ее вывод), если она является предложением, или выясняет невозможность ее вывода, если она не является предложением. Это второй способ задания синтаксиса.

Синтаксис в виде исчисления можно назвать порождающим, или синтезирующим синтаксисом, а синтаксис в виде алгоритма — распознающим, или анализирующим. Любой синтаксис, который может быть полезен, должен допускать оба представления, т. е. описываться разрешимым исчислением. Общие вопросы, относящиеся к моделированию грамматик языков исчислениями и алгоритмами, обсуждались Цейтиным [6].

3. ИСЧИСЛЕНИЯ НЕПОСРЕДСТВЕННО СОСТАВЛЯЮЩИХ

3. 1. Возможные типы задания синтаксиса языка в виде исчисления были исследованы Хомским [1, 2, 3]. Им был определен класс «грамматик фразовых структур», которые задаются исчислениями, имеющими следующий лингвистический смысл. Указываются возможные разложения предложений языка на составляющие их связные непересекающиеся части — конфигурации; для каждой конфигурации указываются ее возможные разложения на составляющие конфигурации низшего порядка и т. д., и наконец для самых «младших» конфигураций указываются составляющие их словоформы. Такое исчисление непосредственных составляющих определяет язык, состоящий из предложений, которые можно построить последовательным развертыванием конфигураций на составляющие кон-

фигурации при использовании перечисленных в исчислении правил. Структурой предложения такого языка является его вывод в исчислении описанного вида, который определяет конфигурационный состав предложения.

3. 2. Например, можно построить грамматику фразовых структур, содержащую следующие правила:

$S \to Q_1N_0V_0$
$Q_1 \to P_1N_1$
$N_1 \to A_1N_1$
$N_0 \to A_0N_0$
$V_0 \to V_0Q_2$
$Q_2 \to P_2N_2$
$N_2 \to A_2N_2$

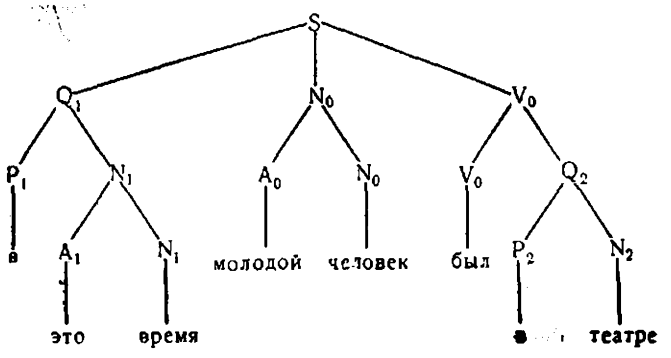
$N_0 \to$ *человек, юноша, ...*
$N_1 \to$ *время, ...*
$N_2 \to$ *театре, доме, ...*
$A_0 \to$ *молодой, высокий, ...*
$A_1 \to$ *это, то, ...*
$A_2 \to$ *хорошем, оперном, городском, ...*
$P_1 \to$ *в, на, за, ...*
$P_2 \to$ *в, на, ...*
$V_0 \to$ *был, ходил, находился, ...*

Здесь латинские буквы с индексами обозначают следующие конфигурации:

S — предложение;
N_0 — группа существительного мужского рода именительного падежа единственного числа;
N_1 — группа существительного среднего рода винительного падежа единственного числа;
N_2 — группа существительного мужского рода предложного падежа единственного числа;
A_0 — группа прилагательного мужского рода именительного падежа единственного числа;
A_1 — группа прилагательного среднего рода винительного падежа единственного числа;
A_2 — группа прилагательного мужского рода предложного падежа единственного числа;
Q_1 — предложная группа с винительным падежом;
Q_2 — предложная группа с предложным падежом;
P_1 — предлог, требующий винительного падежа;
P_2 — предлог, требующий предложного падежа;
V_0 — глагольная группа.

Каждое правило вида $C \to \varphi$ указывает, что конфигурация C может состоять из последовательности конфигураций φ (или из словоформы φ).

В исчислении с такими правилами можно построить вывод русского предложения *в это время молодой человек был в театре* (1) по следующей схеме:



Заключая в скобки каждый отрезок этого предложения, образующий конфигурацию, мы получим следующее представление его конфигурационного состава:

((*в*(*это время*)) (*молодой человек*) (*был*(*в театре*))) (2) (для наглядности здесь и ниже конфигурации, состоящие из единственной словоформы, в скобки не заключены).

3. 3. Синтаксис в виде исчисления непосредственно составляющих представляется довольно естественным, и есть основания предполагать, что живые языки могут быть описаны (хотя бы с достаточным приближением) исчислениями такого типа. Заметим, что эти исчисления разрешимы.

Возможен другой путь построения синтаксиса, также довольно естественный. Можно описывать структуру предложения не в терминах конфигураций, а в терминах грамматических зависимостей между словоформами. Мы рассмотрим общий характер такого синтаксиса.

4. НЕКОТОРЫЕ ОПРЕДЕЛЕНИЯ

4. 1. Пусть $S=a, b, \ldots, c$ — конечная последовательность (цепочка) символов. Занумеруем слева направо символы этой последовательности: $a_1, b_2, \ldots, c_n$, и будем называть символы с такими порядковыми номерами в х о ж д е н и я м и в S.

Можно рассматривать цепочки словоформ, считая каждую словоформу единым символом. Например, в цепочке (1) имеется одно вхождение *молодой*$_4$ словоформы *молодой* и два вхождения $в_1$ и $в_7$ словоформы *в*.

4.2. Пусть для некоторых пар вхождений в S определено отношение $a_i \rightarrow b_j$ («b_j зависит от a_i»), причем:

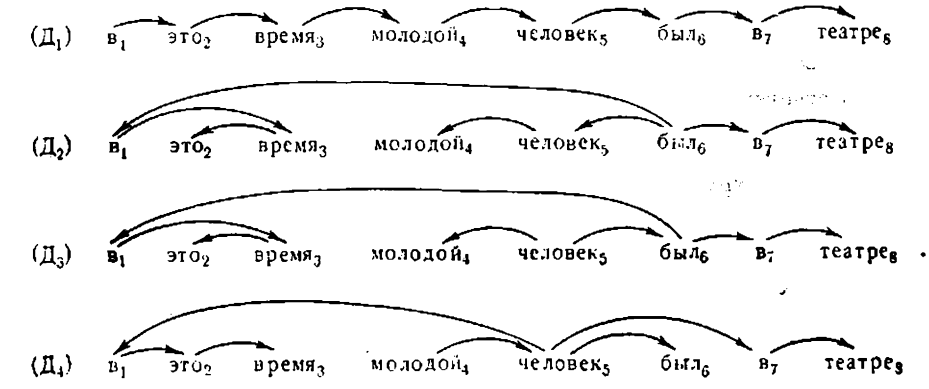
а) имеется одно и только одно вхождение d_k в S, для которого не существует вхождения a_i в S такого, что $a_i \rightarrow d_k$ («вершина»);

б) для каждого вхождения b_j в S, отличного от вершины, имеется одно и только одно вхождение a_i в S ($i \neq j$) такое, что $a_i \rightarrow b_j$;

в) для каждого вхождения a_i в S существует последовательность вхождений $d_k, \ldots, f_l$ такая, что d_k — вершина и $d_k \rightarrow \ldots \rightarrow f_l \rightarrow a_i$.

В этом случае будем говорить, что для S определено д е р е в о з а в и с и м о с т е й. Если $a_i \rightarrow b_j \rightarrow \ldots \rightarrow c_k$ в дереве зависимостей для S, то будем писать $a_i \Rightarrow c_k$ («c_k подчинено a_i»). В частности: $a_i \Rightarrow a_i$; если $a_i \rightarrow b_j$, то $a_i \Rightarrow b_j$; и если d_k — вершина, то $d_k \Rightarrow a_i$ для любого a_i.

Ниже приводится несколько различных деревьев зависимостей для цепочки (1):



4. 3. Дерево зависимостей для цепочки S будем называть к о н ф и г у р а ц и о н н ы м, если выполнено следующее условие:

(А) Для любой четверки натуральных чисел i, j, k, l, если $i < k < j$, $d_l \Rightarrow a_i$ и $d_l \Rightarrow b_j$, то $d_l \Rightarrow c_k$.

Например, деревья $Д_1$ и $Д_2$ для цепочки 1 конфигурационные, а деревья $Д_3$ и $Д_4$ не конфигурационные.

4.4. Если дерево зависимостей Д для S конфигурационно, то для любого вхождения d_1 в S минимальная подцепочка S, содержащая все вхождения, подчиненные d_1, не содержит ни одного вхождения, не подчиненного d_1. Такую подцепочку, определяемую вхождением d_1, будем называть конфигурацией вхождения d_1 относительно дерева Д. Таким образом, условие конфигурационности равносильно условию связности всех конфигураций в цепочке.

Конфигурационное дерево Д определяет для конфигурации каждого вхождения d_1 свое конфигурационное дерево, являющееся поддеревом дерева Д с вершиной d_1.

Конфигурацией вершины конфигурационного дерева для цепочки S является вся цепочка S.

Например, для цепочки (1), заключая в скобки конфигурацию каждого вхождения относительно дерева $Д_1$, получаем:

(3) (в (это (время (молодой (человек (был (в (театре)))))))), а делая то же относительно дерева $Д_2$, получаем:

(4) ((в ((это) время)) ((молодой) человек) был (в (театре))).

4.5. Условие (А) конфигурационности дерева зависимостей имеет более удобную для проверки эквивалентную формулировку:

(Б) Для любой тройки натуральных чисел i, j, k, если $i < k < j$, или $j < k < i$, и имеет место зависимость $a_i \to b_j$, то $a_i \Rightarrow c_k$.

Условие, по существу совпадающее с условием (Б), приводится в [5].

5. СИСТЕМА ЗАВИСИМОСТЕЙ В ЯЗЫКЕ

5.1. Понятие грамматической зависимости между вхождениями словоформ в предложении часто используется при обычном, нестрогом описании синтаксической структуры предложения. Например, для русского языка естественно рассматривать следующие виды зависимостей:

1) существительное, управляемое предлогом, зависит от него;

2) дополнение зависит от глагола, им управляющего;

3) наречие, уточняющее прилагательное, зависит от последнего;

4) прилагательное, определяющее существительное и согласующееся с ним, зависит от него;

5) подлежащее зависит от сказуемого (или наоборот).

Таким образом, отношения управления, согласования и примыкания рассматриваются как частные случаи грамматической зависимости.

5.2. Естественно считать, что в предложении каждое вхождение словоформы зависит только от одного вхождения или же вообще независимо (как сказуемое или подлежащее). Мы, таким образом, приходим к пониманию структуры предложения как иерархической структуры зависимостей, которая является деревом зависимостей при выполнении очевидных требований единственности «главного», независимого вхождения (вершины) и связности всей структуры в целом (пункты «а» и «в» определения дерева зависимостей, см. 4.2).

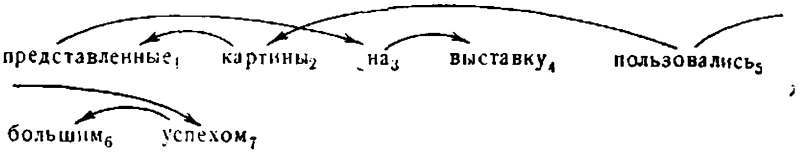
5.3. Определяя структуру предложения в виде дерева зависимостей, нужно определенным образом согласовывать ее с порядком расположения вхождений словоформ в предложении. Наиболее естественным требованием является условие конфигурационности дерева зависимостей, при выполнении которого каждое поддерево структуры предложения укладывается (в линейном расположении вхождений) в виде связной части предложения — конфигурации. Целесообразность этого условия легко проверять на предложениях с интуитивно несомненными зависимостями.

Например, для приведенных в 2. 2 двух предложений можно построить следующие деревья зависимостей:

(5)



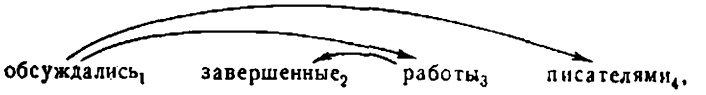
(6)



Не совсем правильный порядок слов, интуитивно ощущаемый в обоих предложениях, находит объяснение в неконфигурационности их деревьев зависимостей. В (5) конфигурация вхождения *завершенные*$_2$ разорвана не относящимся к ней вхождением *работы*$_3$; в (6) — разорвана конфигурация вхождения *представленные*$_1$.

Часто указанием на неконфигурационность дерева зависимостей служит пересечение стрелок в линейной записи дерева, как в (5) и (6). Однако имеются неконфигурационные деревья и без пересечения стрелок (см. дерево $Д_3$ для цепочки (1)).

5. 4. Следует отметить, что первое из приведенных в 5. 3 предложений двусмысленно — можно сопоставить ему другое дерево зависимостей:



которое оказывается конфигурационным, хотя и при таком понимании содержания предложения порядок слов не совсем правильный. Это свидетельствует о том, что одно условие конфигурационности дерева зависимостей не способно достаточно «жестко» определять относительное расположение вхождений словоформ в предложениях языка.

5. 5. При обычном описании структуры двусоставного предложения в терминах зависимостей выделяются две группы вхождений словоформ — группа подлежащего и группа сказуемого, для которых определяются самостоятельные деревья зависимостей. Такой принцип используется и в [7]. Вершины этих деревьев — подлежащее и сказуемое — считаются «главными» вхождениями, и между ними устанавливается симметричная связь. Если мы хотим и в этом случае установить несимметричную, направленную зависимость, то мы должны в первую очередь следить за сохранением конфигурационности дерева. Для предложения (1) естественными являются деревья зависимостей $Д_2$ и $Д_3$. В первом из них подлежащее зависит от сказуемого, и дерево конфигурационно; во втором сказуемое зависит от подлежащего, и дерево неконфигурационно. Последнее объясняется тем, что в нашем предложении группа сказуемого разорвана группой подлежащего. По-видимому, для русского языка этот факт типичен. Очень редки несомненно правильные русские предложения, в которых группа подлежащего разорвана группой сказуемого. Поэтому для русского языка подлежащее следует считать зависящим от сказуемого, если мы хотим пользоваться понятием структуры предложения как конфигурационного дерева зависимостей. Если язык обладает противоположным свойством, то приходится поступать наоборот.

5. 6. *Возможность естественного задания* дерева зависимостей вряд ли вызывает сомнение в случае простых предложений без однородных

членов (хотя и здесь можно привести несколько «сомнительных» примеров).

Однако переходя к сложным предложениям и предложениям с однородными членами, мы наталкиваемся на серьезные трудности.

Например, для предложения *дом, в котором я живу, имеет три этажа* неясно, как можно было бы удовлетворительно определить дерево зависимостей. Очевидно, что следует указать следующие зависимости:



(знаки препинания опускаются и не включаются в структуру предложения). Однако при этом получается два независимых дерева — для главного предложения и для придаточного. Естественно было бы определить зависимость $дом_1 \rightarrow котором_3$ (согласование в роде и числе), но тогда вхождение $котором_3$ будет зависеть от двух вхождений, и мы не получим дерева зависимостей.

Подобные трудности возникают и при рассмотрении предложений с однородными членами. Достаточно рассмотреть предложение *юноши и девушки пели песни и танцевали.*

5. 7. Сказанное наводит на мысль, что полезно было бы расширить понятие структуры предложения, определяя для них не только деревья зависимостей, но иногда и графы зависимостей, или же вводя кроме обычных зависимостей в деревьях дополнительные связи другого характера между вхождениями или целыми конфигурациями вхождений. Мы этого делать здесь не будем. Нашей задачей является исследование характера аксиоматического описания синтаксиса, основанного на понятии структуры предложения как конфигурационного дерева зависимостей. Ниже мы вернемся к вопросу о достаточности такого понимания структуры предложения.

6. ЗАВИСИМОСТИ И ВАЛЕНТНОСТИ

6. 1. Мы рассмотрели содержательный смысл установления структуры предложения как конфигурационного дерева зависимостей. Перейдем теперь к выяснению возможного вида аксиоматического описания синтаксиса, которое использует это понимание структуры.

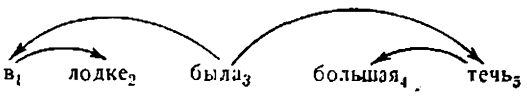
При данном конечном перечислении словоформ можно попытаться указать для каждой словоформы список всех словоформ, которые могут зависеть от вхождения этой словоформы в предложениях языка. Другими словами, мы аксиоматически задаем «допустимые зависимые контексты» каждой словоформы. После этого предложение определяется как такая цепочка словоформ, для которой можно построить конфигурационное дерево зависимостей, в котором каждая зависимость допустима (д о п у с т и м а я с т р у к т у р а).

Указание на возможную зависимость вхождения словоформы b от вхождения словоформы a в предложении языка мы будем называть в а л е н т н о с т ь ю словоформы a на словоформу b и записывать его как упорядоченную пару (a, b). Таким образом, в простейшем случае с и н т а к с и с з а в и с и м о с т е й представляет собой конечное перечисление валентностей в виде упорядоченных пар словоформ.

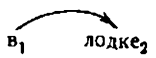
6. 2. Например, при задании естественного синтаксиса зависимостей в указанном смысле для русского языка мы в число валентностей включаем следующие пары:

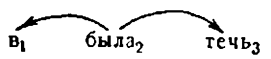
(7) (в, лодке), (течь, большая), (была, течь).

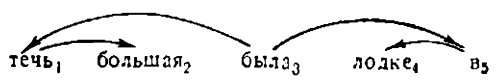
Тогда предложением будет являться цепочка *в лодке была большая течь* с допустимой структурой

(8) 

Однако при наличии в синтаксисе валентностей (7) предложениями будут и цепочки *в лодке*, *в была течь*, *течь большая была лодке в*, и т. п., так как для них можно указать следующие допустимые структуры:

(9) 

(10) 

(11) 

Если же в синтаксис, кроме (7), будут включены также и валентности (*начала*, *лодка*), (*начала*, *течь*) и (*в*, *шлюпке*), то в число предложений попадут и цепочки *лодка начала большая течь* и *лодке в шлюпке лодке была течь* с допустимыми структурами

(12) 

(13) 

6. 3. Таким образом, синтаксис указанного вида всегда будет относить к предложениям заведомо «неправильные» цепочки или цепочки, не образующие с обычной точки зрения целого предложения. Он будет слишком свободным.

Выясним, какие грамматические явления не учитываются таким синтаксисом.

1. Не всякая словоформа может входить в предложение как вершина допустимого дерева зависимостей (см. (9)) — ограничение на вершины деревьев.

2. Некоторые словоформы не могут входить в предложение, относительно допустимых деревьев которых они не имеют зависимых словоформ (например, предлог в (10)) — «обязательность» реализации некоторых валентностей.

3. Некоторые пары вхождений словоформ, связанных в допустимой структуре предложения зависимостью, могут располагаться только в определенном порядке относительно друг друга (например, пара «предлог — существительное» в (11)) — ограничение на относительные позиции вхождений, связанных зависимостью.

4. Не всякие валентности могут одновременно реализоваться в виде зависимостей в одной и той же допустимой структуре предложения (например, в (12) одновременно присутствуют зависимости *начала*$_2$ → *течь*$_4$ и *течь*$_4$ → *большая*$_3$, которые несовместны, ввиду омонимии словоформы

течь) — несовместность валентностей, объясняющаяся грамматической омонимией словоформ.

5. Некоторые валентности могут реализоваться в допустимой структуре предложения только по одному разу для каждого вхождения соответствующей словоформы (см. (13)) — «однократность» и «многократность» валентностей.

Мы ограничимся этими фактами, хотя можно было бы и продолжить перечисление недостатков рассматриваемого синтаксиса.

6. 4. Построение более удовлетворительной схемы синтаксиса мы начнем с учета наиболее важного свойства естественных языков — «грамматической омонимии» словоформ. Это свойство содержательно проявляется в несовместности некоторых валентностей. Оно состоит в том, что у некоторых словоформ имеется несколько «грамматических значений». Для аксиоматического задания синтаксиса нет нужды как-то определять термин «грамматическое значение» — нам важен не смысл этого термина, а способ различения «омонимии».

По некоторым, иногда очевидным соображениям, мы можем «омонимичную» словоформу расщепить на несколько «грамматических значений». Перенумеровав для каждой словоформы эти «значения», будем в дальнейшем рассматривать не саму словоформу, а словоформу с указанным номером ее «грамматического значения». Так, например, вместо словоформ *течь* и *начала* мы можем ввести расщепления *течь*[1] (существительное), *течь*[2] (глагол), *начала* [1] (глагол), *начала* [2] (существительное единственного числа в родительном падеже), *начала* [3] (существительное множественного числа именительного падежа). После этого зададим валентности не просто для словоформ, а для словоформ с указанием номера «грамматического значения»: (*начала* [1], *течь* [2]), (*течь* [1], *большая*), (*начала* [2], *большого*), (*начала* [3], *необходимые*), и т. п.

6. 5. Возможен другой, более удобный и компактный учет «грамматической омонимии» в валентностях. Введем список «абстрактных грамматических значений» — некоторых вспомогательных символов A, B,. . ., L, которые будем называть к л а с с а м и. Каждой словоформе соотнесем некоторое ненулевое количество классов. Если словоформе a соотнесено несколько классов A, C,. . ., E, то мы говорим, что a имеет «грамматические значения» $A, C,. . ., E$; если словоформе соотнесен только один класс, то мы говорим, что она лишена «омонимии».

Можно интерпретировать классы как символы дистрибутивных классов и применять при построении такого синтаксиса методы дистрибутивного анализа. Соотнесение класса A словоформе a в этом случае понимается как указание на принадлежность a к дистрибутивному классу A.

Вместо валентностей словоформ вводятся валентности классов — упорядоченные пары (A, B). Определение предложений языка, описываемого заданным списком валентностей классов, происходит в два этапа. Рассматриваются всевозможные цепочки классов. Цепочка классов называется допустимой структурой, если для нее можно определить конфигурационное дерево зависимостей (так же, как и для словоформ), в котором каждая зависимость будет реализовать имеющуюся в синтаксисе валентность соответствующего класса. Цепочка словоформ называется предложением, если для каждого вхождения словоформы можно выбрать один из классов, соотнесенных этой словоформе, так, что полученная цепочка классов будет допустимой структурой.

Приведенное определение не претендует на строгость — оно лишь поясняет характер наших построений. Точное определение синтаксиса и предложений с учетом «омонимии» заняло бы много места.

6. 6. Взяв за основу представление синтаксиса в виде конечного перечисления валентностей классов, постараемся теперь учесть остальные свойства языков, перечисленные в 6. 3, с точки зрения зависимостей.

Относительные позиции вхождений, связанных зависимостью в допустимых структурах, можно учесть, снабжая валентности классов знаками относительных позиций + или —, указывающими на направленность зависимости, реализующей валентность (соответственно вправо или влево). Одна и та же валентность может присутствовать в списке с обоими знаками — это значит, что относительные позиции вхождений словоформ несущественны для реализации этой валентности. Так, валентность предлога на управляемое им существительное будет иметь вид $+(A, B)$, валентность существительного на согласующееся с ним прилагательное — вид $-(A, B)$, а валентность глагола на предлог, как правило, вид $+(A, B)$.

«Обязательность» и «необязательность» реализации валентности будем учитывать, снабжая валентности знаком «!» («обязательность») или «?» («необязательность»). При этом у одной и той же валентности не допускается указание обоих знаков одновременно; каждая валентность синтаксиса снабжается одним и только одним из этих двух знаков. Так, валентность предлога на существительное будет иметь вид $+(A, B)!$, а валентность существительного на прилагательное — вид $-(A, B)?$.

Можно учитывать и кратность валентности. При этом надо различать k-кратные валентности, которые могут реализовываться для одного и того же вхождения не более чем k зависимостями ($k=1, 2, \ldots$), и «бесконечно кратные» валентности, количество реализаций которых для каждого вхождения не ограничено никаким фиксированным числом. Введение «бесконечно кратных» валентностей может облегчить задание синтаксиса ценой допущения сколь угодно большого количества зависимостей, исходящих из одного и того же вхождения в допустимой структуре.

Наконец, все классы, используемые в синтаксисе, можно разделить на классы, которые могут стоять в вершине допустимой структуры, и классы, которые не могут стоять в этой позиции.

Достаточно подробная лингвистическая классификация типов валентностей принадлежит Б. М. Лейкиной [8].

7. ИСЧИСЛЕНИЯ ВАЛЕНТНОСТЕЙ

7. 1. В общем случае исчисление валентностей представляет собой конечный аппарат описанного типа, определяющий язык структур — множество цепочек классов в заданном алфавите, являющихся допустимыми структурами (см. 6. 5). Если каждому классу сопоставлены некоторые словоформы (символы другого алфавита), то каждой структуре из языка структур очевидным образом соотносятся определенные цепочки словоформ.

Пусть H — исчисление валентностей, f — отображение, сопоставляющее каждой словоформе a непустое множество классов $f(a)$. Через T_H будем обозначать язык структур, определяемый исчислением H, а через $L_{H,f}$ — язык, состоящий из всех цепочек словоформ $S = ab\ldots c$, для которых существует структура $S — AC\ldots D$ из T_H такая, что $A \in f(a)$, $C \in f(b), \ldots, D \in f(c)$.

Отображение f, сопоставляемое исчислению валентностей (оно указывает для каждой словоформы ее «грамматические значения»), назовем словарем. Одно и то же исчисление валентностей с разными словарями может определять различные языки $L_{H,f}$. Исчисление H с заданным словарем f будем называть синтаксисом зависимостей $[H, f]$.

Можно определять различные типы исчислений валентностей и синтаксисов зависимостей, используя средства характеристики валентностей из 6. 6 или вводя другие средства. Различные возможные формулировки таких исчислений обсуждались в работах Белецкого [9] и автора [10].

7. 2. Если в исчислении непосредственно составляющих при определении процесса построения предложений используются правила только одного типа (подстановка цепочки символов вместо вхождения одиночного символа), то в исчислении валентностей могут использоваться правила различных типов, что делает синтаксическое описание более содержательным и богатым. Может оказаться, что некоторые средства характеристики валентностей можно заменить комбинацией других средств, т. е. что различные средства характеристики валентностей в этом смысле не независимы. С первого взгляда не очевидно, каким средствам отдавать предпочтение.

Представляет интерес выяснение минимального и достаточно простого набора средств характеристики валентностей, к которому можно было бы сводить другие наборы средств.

Поскольку заранее нельзя предусмотреть все возможные средства характеристики валентностей, поставленная задача требует уточнения.

7. 3. Любой язык можно пытаться описывать либо исчислением непосредственно составляющих, либо синтаксисом зависимостей. Можно показать, что если синтаксис зависимостей $[H, f]$ использует средства характеристики валентностей, перечисленные в 6. 6, то язык $L_{H,f}$, им определяемый, можно задать и исчислением непосредственно составляющих.

Таким образом, изучая языки, определяемые синтаксисом зависимостей с естественными средствами характеристики валентностей, мы не выходим за рамки языков со структурой непосредственно составляющих. Разумно предполагать, что естественные языки обладают структурой непосредственно составляющих. Ввиду этого нашу задачу можно переформулировать следующим образом: найти минимальный и достаточно простой набор средств характеристики валентностей, при использовании которого можно было бы описать синтаксисом зависимостей любой язык со структурой непосредственно составляющих. Для решения этой задачи мы можем применять точные математические средства.

Заметим, что если поставленная задача имеет решение, то понятие структуры предложения как конфигурационного дерева зависимостей теоретически оказывается достаточным для аксиоматического задания синтаксиса естественных языков (см. 5. 6 и 5. 7).

7. 4. При изучении различных типов задания валентностей необходимо следить за тем, чтобы рассматриваемые исчисления валентностей были разрешимыми (см. 2. 6). Исчисления валентностей, которые используют средства, перечисленные в 6.6, разрешимы. Характер разрешающих алгоритмов для таких исчислений описан в [5] и [11]. Эти алгоритмы существенно используют конфигурационность деревьев зависимостей.

8. ЗАКЛЮЧЕНИЕ

8. 1. Задача аксиоматического описания языков средствами исчислений впервые в явном виде была сформулирована Хомским [1, 2]. Предпосылки такого описания содержатся в методах структурной лингвистики и наиболее ярко проявляются в работах американской школы дескриптивной лингвистики [12, 13, 14]. Методы дескриптивного анализа имеют характер правил, с помощью которых выявляются элементы возможного аксиоматического аппарата для описания языка. Результатом применения этих правил к ограниченному объему исходного материала является некоторая абстрактная структурная схема. Ее состоятельность проверяется на неучтенных языковых явлениях по тем или иным критериям.

Таким образом, индуктивные методы служат для построения дедуктивного (аксиоматического) описания.

8. 2. Мы рассмотрели только те методы аксиоматического описания языка, которые относятся к синтаксису. Морфология предполагалась заданной в виде простого списка словоформ.

Более содержательным было бы построение нетривиальной морфологии, описывающей процессы словоизменения и словообразования на базе некоторого исчисления. Основной проблемой здесь является внутреннее, структурное определение понятия слова. Пути решения этой проблемы пока неясны, хотя можно предположить, что наиболее естественным явилось бы понимание слова как некоторого множества словоформ. Возможно, что структурное определение слова будет использовать и элементы синтаксического описания.

8. 3. Кроме рассмотренных исчислений непосредственных составляющих и валентностей, при описании языка целесообразно использовать и трансформационные методы. Как было указано Хомским [1, 2], синтаксическое описание может быть существенно упрощено, если задавать сперва исчислением только ограниченное количество «базисных» предложений, определяя затем для таких предложений (с указанными структурами) дополнительное исчисление трансформаций, строящих новые предложения из «базисных».

Трансформационные методы внутреннего описания языка могут быть связаны со следующим пониманием семантики. Для некоторых пар предложений языка определяется отношение эквивалентности. Это отношение задает классы эквивалентных предложений, которые можно рассматривать как совокупности предложений с общими семантическими свойствами. Такой подход к семантике можно считать частным случаем понимания семантики как «семантики соответствия» (см. 1. 2 и 1. 3), когда соответствие устанавливается между двумя совпадающими языками.

Отношение эквивалентности можно устанавливать на основе трансформаций. Если одно предложение может трансформироваться в другое (при использовании заданных операций) и наоборот, то эти два предложения считаются эквивалентными. Таким образом, трансформации в этом понимании представляют собой преобразования, инвариантом которых является содержание сообщения. «Семантика эквивалентности» определяется указанным набором трансформационных правил.

ЛИТЕРАТУРА

1. N. Chomsky. Three models for the description of language. «IRE Transactions on Information Theory», IT-2, №. 3, 1956.
2. N. Chomsky. Syntactic Structures. The Hague, 1957.
3. N. Chomsky. On certain formal properties of grammars. «Information and Control», 2, №. 1, 1959.
4. N. Chomsky and G. A. Miller. Finite state languages. «Information and Control», 1, 1958.
5. K. E. Harper and D. G. Hays. The use of machines in the construction of a grammar and computer program for structural analysis. «Proceedings of the International Congress on Information Processing», UNESCO, Paris, 1959.
6. Г. С. Цейтин. К вопросу о построении математических моделей языка. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста». М., изд. ВИНИТИ, 1961.
7. И. И. Ревзин. Формальный и семантический анализ синтаксических связей в языке. Сб. «Применение логики в науке и технике». М., Изд-во АН СССР, 1960.
8. Б. М. Лейкина. Некоторые аспекты характеристики валентностей. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста». М., изд. ВИНИТИ, 1961.
9. М. И. Белецкий. Несколько моделей языка. Доклад на конференции по обработке информации, машинному переводу и автоматическому чтению текста (неопублик.).
10. С. Я. Фитиалов. О моделировании синтаксиса для машинного перевода. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста». М., изд. ВИНИТИ, 1961.

11. Г. С. Цейтин, Л. Н. Засорина. О выделении конфигураций в русском предложении. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста». М., изд. ВИНИТИ, 1961.
12. Z. Harris. Methods in Structural Linguistics. Chicago, 1951.
13. Ch. Hockett. A Course in Modern Linguistics. New York, 1958.
14. Г. Глисон. Введение в дескриптивную лингвистику. М., ИЛ, 1959.

ON SYNTAX MODELLING IN STRUCTURAL LINGUISTICS

Summary

There are two types of linguistic description — «external» or semantic, and «internal» or grammatical. The internal syntactic description or formal syntax of a language is a calculus generating a set of strings of symbols — the set of correct sentences of the language described. The valency calculus based on the conception of the relation of dependance between words in a sentence is one of the most natural kinds of syntactic calculus. Different types of the valency calculus are considered as a possible basis for the construction of formal models for grammars of languages. The connection between valency calculus and «phrase structure» calculus is discussed.

К. И. Бабицкий

К ВОПРОСУ О МОДЕЛИРОВАНИИ СТРУКТУРЫ ПРОСТОГО ПРЕДЛОЖЕНИЯ [1]

Предлагаемая работа посвящена изучению структуры простого предложения в немецком языке на материале математических текстов. Таким материалом послужил полугодовой (№№ 1—6) комплект журнала ZAMM («Zeitschrift fur angewandte Mathematik und Mechanik») за 1958 г., содержащий около 2000 независимых предложений. Элементарный прием, с которого начиналось исследование, состоял в замене последовательности слов предложения последовательностью символов и в попытках классифицировать полученные цепочки. Целью работы было уложить эти цепочки в некоторую таблицу или другую формальную систему, которая и явилась бы искомой моделью грамматики немецкого простого предложения. В принципе, чем меньше была бы такая таблица, тем лучше, ибо это означало бы, что мы сумели выявить наиболее общие свойства изучаемых структур. Практически число отображающих цепочек должно быть обозримым, скажем, лежащим в пределах сотни.

Этому требованию удовлетворяет следующий порядок отображения множества независимых предложений в множество цепочек символов.

Берем предложение и опускаем в нем все, кроме подлежащего, сказуемого и дополнения или дополнений. При наличии однородных членов опускаем в каждом ряду все, кроме одного. Опускаем также обособленные обороты и члены предложения. Подлежащее заменяем символом S, дополнения — символами O_1 (прямое) и O_2 (косвенное). При этом артикль и предлог не отражаются — они считаются составляющими одно целое с существительным. Словесный состав сказуемого отражается полностью (за исключением того, что отделяемая приставка считается составляющей одно целое с личной формой глагола). Глагол заменяем символом Z и различаем следующие формы глагола: Л (Z) — личные формы, И (Z) — инфинитив, $П_2$ (Z) — причастие второе. Именную часть сказуемого заменяем символом Р. Заменив соответствующими символами указанные члены предложения и их части, расставляем символы в том порядке, в котором стоят в предложении соответствующие слова или сочетания слов, и соединяем их знаком ⌢.

Выражение X⌢Y означает, что элемент X отображаемого множества предшествует элементу Y. Построенные таким способом цепочки называем образами предложений. Ниже показаны примеры построения обра-

[1] Автор благодарит И. И. Ревзина за непосредственное и плодотворное участие в определении целей, отправных положений и методики работы, а также своих товарищей по группе машинного перевода ИТМ и ВТ С. С. Белокриницкую и В. В. Паршина за ряд ценных советов и замечаний. Особенно полезной была помощь Н. М. Нагорного в разработке формальной стороны работы.

зов заданных предложений. (Подлежащее набрано полужирным шрифтом, сказуемое — курсивом, дополнение — вразрядку.)

1) E i n e n breiten R a u m *nehmen* auch **die** vorbeiliegenden **Forschungsarbeiten** *ein*:

$$O_1\hat{}Л\,(Z)\hat{}S$$

2) **Diese Methode** *wurde* a u f f o l g e n d e S c h w i n g u n g s s y s t e m e *ausgedehnt*:

$$S\hat{}Л\,(Z)\hat{}O_2\hat{}П_2\,(Z)$$

3) Im Schreibwerk *läuft* in Abszissenrichtung **ein Papierstreifen**:

$$Л\,(Z)\hat{}S$$

4) *Gross ist* **der Kreis** von Untersuchungen *gewesen*:

$$P\hat{}Л\,(Z)\hat{}S\hat{}П_2(Z)$$

Наконец, один из самых сложных случаев, встреченных в нашем тексте.

5) In diesem Fall *kann* **das Maximum** *verkleinert worden sein*:

$$Л\,(Z)\hat{}S\hat{}П_2\,(Z)\hat{}П_2(Z)\hat{}И\,(Z)$$

При попытке разобраться в строении множества образов можно заметить прежде всего следующий основной факт. Среди образов имеется небольшое число элементарных, простых по своей структуре, по отношению к которым всякий другой, самый сложный образ можно считать производным. Если отнести к числу элементарных, например, следующие образы:

$S\hat{}Л\,(Z)$	(der Vogel fliegt)
$S\hat{}Л\,(Z)\hat{}O_1$	(der Mensch baut das Haus)
$S\hat{}Л\,(Z)\hat{}П_2\,(Z)$	(das Haus wird gebaut)
$S\hat{}Л\,(Z)\hat{}P$	(ich bin Doktor)

то образы, приведенные в примерах, можно рассматривать как полученные из элементарных образов посредством некоторых регулярных преобразований. Проследим, как это делается.

Пример (3):

$S\hat{}Л\,(Z)$	Ein Papierstreifen läuft
$Л\,(Z)\hat{}S$	Läuft ein Papierstreifen.

Преобразование здесь состояло в том, что мы поменяли местами подлежащее и простое сказуемое, т. е. глагол в личной форме.

Пример (1):

$S\hat{}Л\,(Z)\hat{}O_1$	Die Forschungsarbeiten nehmen einen Raum ein
$O_1\hat{}Л\,(Z)\hat{}S$	Einen Raum nehmen die Forschungsarbeiten ein

Здесь мы поменяли местами подлежащее и прямое дополнение.

Пример (4):

$S\hat{}Л\,(Z)\hat{}P$	Der Kreis ist gross
$S\hat{}Л\,(Z)\hat{}P\hat{}П_2\,(Z)$	Der Kreis ist gross gewesen
$P\hat{}Л\,(Z)\hat{}S\hat{}П_2(Z)$	Gross ist der Kreis gewesen

В данном случае мы заменили глагол, стоявший в личной форме, той же формой вспомогательного глагола, а замененный глагол в форме Partizip II поставили в конце предложения; после этого мы поменяли местами подлежащее и именную часть сказуемого. По случайности, заменяемый и заменяющий глаголы оказались одним и тем же глаголом *sein*. В факте замены, однако, легко убедиться на любом примере образования перфекта из настоящего времени.

В данном же примере наглядно выступает тонкое различие в употреблении глагола *sein*.

Пример (5):

S^Л (Z)^$П_2$ (Z)	Das Maximum wird verkleinert
S^Л (Z)^$П_2$ (Z)^$П_2$ (Z)	Das Maximum ist verkleinert worden
S^Л (Z)^$П_2$ (Z)^$П_2$ (Z)^И (Z)	Das Maximum kann verkleinert worden sein
Л (Z)^S^$П_2$ (Z)^$П_2$ (Z)^И (Z)	Kann das Maximum verkleinert worden sein.

В этом примере мы, во-первых, заменили глагол, стоявший в личной форме, той же формой глагола *sein*, а замененный глагол поставили в форме Partizip II в конце цепочки; во-вторых, в полученной цепочке заменили глагол в личной форме той же формой глагола *können*, а замененный глагол поставили в форме инфинитива в конце цепочки; в-третьих, в полученной цепочке поменяли местами подлежащее и глагол в личной форме.

Из приведенных примеров видно, что множество образов можно разбить на подмножества, каждое из которых включает некоторый элементарный образ и все образы производные от него; внутри данного подмножества образы различаются только тем, какие преобразования, в каком числе и порядке были приложены к элементарному образу для их получения.

Сами преобразования могут быть двух видов: перестановки и трансформации. В связи с этим структура производного образа отчетливо распадается на две части: «левую» и «правую» (см. пример 5). «Правая» часть состоит из неличных форм глагола, количество и качество которых отражает историю трансформаций, произведенных в элементарном образе. «Левая» часть повторяет (с учетом перестановок) структуру самого элементарного образа. То, что перестановки затрагивают только «левую» часть, верно лишь приблизительно, т. е. для подавляющего числа случаев. В некоторых случаях это положение не соблюдается. Примером может служить предложение Gegeben sei folgendes System der Differenzialgleichungen, имеющее образ: $П_2$ (Z)^Л(Z)^S. Поэтому слова «левая» и «правая» поставлены в кавычки. Интересно, однако, что в то время, как перестановки символов «левой» части могут быть самыми разнообразными (особенно в элементарных образах более сложного вида — см. ниже), перестановки символов «правой» части имеют всегда характер «кольцевого переноса», при котором порядок следования символов не нарушается.

Таким образом, искомая модель «грамматики» немецкого простого предложения (в том ограниченном смысле, который определен указанным выше порядком построения образов) должна включать набор элементарных образов и набор преобразований, а также, возможно, и некоторые другие закономерности. Подобная модель охватывает, следовательно, лишь два круга явлений грамматического строя простого предложения: взаимоотношения между подлежащим, личной формой глагола, дополнениями и именной частью составного сказуемого и взаимоотношения между личной и неличными формами глаголов в составе сложного сказуемого, т. е. образование аналитических временных, модальных и других форм глагола.

Выбор элементарных образов (которые в дальнейшем будем называть аксиомами) и операторов преобразования представляет собой наиболее трудную проблему, от успешного решения которой зависит сама возможность создания формальной, т. е. автоматически действующей модели.

Подход к решению (которое пока еще нельзя считать вполне удовлетворительным) осуществлялся в настоящей работе по методу последовательного приближения. На небольшом отрывке текста (около 100 предложений) были получены образы; из них выделены аксиомы и определены преобразования. Полученная модель проверена на образах, выведенных из большего отрывка (350 предложений). Модель усовершенствована с учетом обнаружившихся противоречий. Измененная модель вновь проверена на образах, выделенных из текста объемом в 2000 предложений и вновь переделана с учетом новых фактов и накопившегося опыта. Этот третий (и, по-видимому, не окончательный) вариант излагается ниже.

Не останавливаясь на указанных этапах, необходимо заметить следующее. Руководящим соображением при проверке и корректировании модели было требование ее обратимости. Это значит, что, взяв простое предложение, построив его образ, запомнив (скажем, занумеровав) лексические значения входящих в образ слов и возведя путем последовательных преобразований данный образ к аксиоме, нужно иметь возможность путем приложения к аксиоме в обратном порядке обратных преобразований получить исходный образ, а по нему (подставив лексические значения) — и само исходное предложение, точнее тот его «препарат», который останется, если опустим определения, обстоятельства, однородные и обособленные члены. Ввиду чрезвычайной сложности такой задачи практически преследовалась обратимость в пределах анализа — синтеза от производного образа к аксиоме и обратно к производному образу. При этом в качестве первостепенной проблемы выдвинулся вопрос классификации глаголов.

Важность различения лексико-грамматических групп глаголов для понимания структуры образов видна из следующих двух примеров (здесь и ниже в качестве примеров приводятся, для простоты, «препараты» предложений).

Пример (6): die Ergebnissen werden genannt; die Kurve ist gegangen.

Эти предложения имеют один и тот же образ $S\hat{}Л(Z)\hat{}П_2(Z)$. И только указав, что в первом случае в личной форме стоит глагол *werden*, а во втором — *sein* (при непереходности глагола в Partizip II), можно увидеть (из образа), что предложения эти синтаксически различны.

Пример (7): Предложения типа: die Reihe muss konwergieren; wir werden vorgehen также имеют один и тот же образ $S\hat{}Л(Z)\hat{}И(Z)$, и их синтаксическая разнородность выясняется только при указании на характер глаголов в личной форме.

Разбив все глаголы на лексико-грамматические группы и дополнив образы указанием, к каким группам относятся входящие в них глаголы, мы устраним указанную структурную «омонимию» образов. Гораздо более важной представляется, однако, обратная связь: определение лексико-грамматической принадлежности глаголов в зависимости от структуры образов, в которые они входят. Эту связь мы уже наблюдали в примере (4), где глагол *sein* имел различное значение в зависимости от того, входил ли он в аксиому или участвовал, в качестве заменяющего, в преобразовании.

Поскольку наиболее интересные грамматически так называемые служебные глаголы, как правило, весьма многозначны, мы получаем в этой связи объективный метод дифференцирования их значений, позволяющий в некоторых случаях говорить о настоящей грамматической омонимии.

В полной мере этот вопрос раскрывается лишь в изложении самой модели, к которому мы и переходим.

Если рассмотреть множество образов, можно обнаружить, что один образ от другого может отличаться 8 различными способами. Эти соотношения между образами, отражающие соотношения между предложениями, следующие.

I. S^Л (Z)	Die Gleichung gilt
S^Л (Z)^O_1	Die Gleichung hat eine Lösung

Это соотношение между предложением с непереходным глаголом (которое не содержит прямого дополнения) и предложением с переходным глаголом (содержащим прямое дополнение).

II. S^Л (Z)	Die Funktion läuft
S^Л (Z)^O_2	Die Funktion genügt einer Bedingung

Здесь соотносятся отсутствие и наличие косвенного дополнения и, соответственно, непереходный и косвенно-переходный глаголы.

III. S^Л (Z)^O_1	Wir schätzen den Vektor ab
S^Л (Z)^O_1^P	Wir setzen die Elementen als reel voraus

Это соотношение между предложениями глагольным и именным. Второе отличается от первого наличием предикатива и изменением грамматической функции глагола.

IV. S^Л (Z)^O_1	Die Entwicklung hat die Form
S^Л (Z) O_1^П (Z)	Die Entwicklung muss die Form haben

Здесь нижнее предложение получается из верхнего заменой глагола в личной форме модальным глаголом в той же форме и постановкой замененного глагола в форме инфинитива в конце предложения.

V. S^Л (Z)	G verschwindet
S^Л (Z)^$П_2$ (Z)	G ist verschwunden

Второе предложение получаем из первого, заменив глагол в личной форме той же формой вспомогательного и поставив замененный глагол в форме Partizip II в конце предложения.

VI. S^Л (Z)^P	Das Wert ist positiv
S^Л (Z)^P^П (Z)	Das Wert wird positiv sein

Второе предложение получается из первого, если глагол в личной форме заменим вспомогательным глаголом в той же форме, а инфинитив от связочного глагола поставим в конце предложения.

Соотношения I—III, с одной стороны, и IV—VI, с другой, представляют собой два резко различных типа. Помимо различия структуры (которое не так уж велико) они отличаются в основном тем, что соотношения второго типа могут не менять (как в наших примерах) предметного значения предложения, а соотношения первого типа суть связи между вещественно разными предложениями (хотя бы подлежащее и дополнение, например в случае III, были лексически одинаковы). В первом случае мы имеем соответствия, во втором — преобразования, и ясно, что соответствия — это то, чем отличаются друг от друга аксиомы, а посредством преобразований мы можем получать из аксиом производные образы, а затем (в силу

неизменности предметного значения) получать из производных образов исходные аксиомы с помощью обратных преобразований.

Таким образом будет удовлетворяться требование обратимости. К преобразованиям относится также соотношение следующего типа.

VII. S^Л(Z)^P Der Kreis ist gross
P^Л(Z)^S Gross ist der Kreis

Преобразования IV—VI называются трансформациями, преобразование VII — перестановкой. Не останавливаясь на различных видах перестановок и их синтаксическом значении (вопрос интересный, но достаточно простой), будем считать образы — и, соответственно, предложения — относящимися к одному типу, если они различаются только порядком слов, т. е. будем рассматривать в качестве элементов подмножеств, о которых шла речь выше, классы образов, отличающихся только порядком входящих в них символов. Наконец, имеется еще одно соотношение, более сложного типа.

VIII. S^Л(Z)^O_1 Der Verfasser gibt ein Schema
[S^Л(Z)^$П_2$(Z) Ein Schema wird gegeben

Элементарные действия, из которых складывается это соотношение, распадаются на две части. Первая часть полностью соответствует трансформациям IV—VI: мы заменяем глагол, стоящий в личной форме, той же формой вспомогательного глагола, а замененный глагол ставим в конце в форме Partizip II. Вторая часть формально проста: опускается прямое дополнение. В этом смысле вторая часть подходит под тип соответствий I—III (наличие—отсутствие). Но если говорить о предметном значении, то фактически происходит перестановка слов, выражающих субъект и объект действия, после чего эти слова обмениваются друг с другом своими формами подлежащего и прямого дополнения, и, наконец, слово-субъект опускается.

Возникает вопрос — самый трудный во всей работе: чем считать соотношение VIII — соответствием или преобразованием? В качестве преобразования оно, очевидно, необратимо; как ни считать, предметно или синтаксически, одно слово все равно исчезает и обратное преобразование его не восстанавливает. Но если считать его соответствием, то мы получим аксиому с вспомогательным глаголом и неличной формой глагола. При этом сильно ухудшаются шансы на то, что нам удастся создать систему объективной классификации глаголов. Таково противоречие, которое необходимо разрешить, если мы хотим получить модель, хотя бы и несовершенную.

Исходя из того, что строгая обратимость — слишком хорошая вещь, чтобы быть фактом, считаем соотношение VIII трансформацией. Решение, очевидно, может быть и иным, поэтому не будем забывать, что мы совершили акт произвола, разрезав эту естественную пуповину, соединяющую преобразования с аксиомами.

Но как только этот болезненный акт совершен, мы получаем восемь аксиом, связанных между собой соответствиями I—III, и четыре трансформации: IV, V, VI и VIII (перестановки мы согласились пока не рассматривать). Обозначив соответствия символами: I—D, II—Д, III—İ и трансформации символами: VIII—З, IV—M, V—B, VI—F — получаем следующую классификацию глаголов. Глаголы делятся на знаменательные (образующие аксиомы) и вспомогательные (входящие в трансформации в качестве заменяющих). Знаменательные глаголы имеют нижним индексом трехзначное двоичное число dдi, где символы означают: d — прямая переходность (соотв. D), д — косвенная переходность (соотв. Д), i —

связочность (соотв. i) и могут принимать значения: 0 — отсутствие признака, 1 — наличие признака.

Таким образом, символы 8 групп знаменательных глаголов (по числу аксиом) следующие:

Z_{000} — непереходные
Z_{100} — прямо-переходные
Z_{010} — косвенно-переходные
Z_{110} — двойного управления
Z_{001} — } то же, но с семантическим множителем
Z_{101} — } связочности
Z_{011} — }
Z_{111} — }

Отсутствие трехзначного нижнего индекса означает, что глагол не образует аксиом.

Вспомогательные глаголы делятся на инфинитивные и причастные в зависимости от того, какой неличной формы глагола они требуют. Если глагол входит в трансформацию (является вспомогательным), он имеет нижним индексом одноразрядное двоичное число, где 0 означает инфинитивность, 1 — причастность.

В частности, глагол может принадлежать только к одной из этих 16 групп, но, вообще говоря, эти группы пересекаются, т. е. одна и та же глагольная лексема в разных контекстах может выполнять функции, свойственные нескольким различным группам. При этом о функции знаменательного глагола мы судим формально на основании наличия—отсутствия O_1, O_2, и P, а о функции вспомогательного — по наличию $П_2$ или И. Многозначным глаголам соответствует несколько кодов, по числу выполняемых функций. Для окончательного варианта модели необходимо иметь словарь глаголов с указанием (в кодовом обозначении) всех функций каждого глагола. На данном этапе эта работа лишь намечена, т. е. есть небольшой словарь, охватывающий всего около сотни наиболее интересных глаголов, причем без претензий на точность определения значений. Из этого словаря (он дается в приложении I) можно, однако, наглядно уяснить практическое осуществление принципов классификации глаголов и кодирования грамматической информации о них (в рамках данной модели).

Следует иметь в виду, что многозначность глаголов указывается в основном лишь в пределах значений, встреченных в анализируемом тексте.

Выпишем из словаря все глаголы, выполняющие функцию вспомогательных, и сгруппируем их. При этом многозначность, обозначенную в словаре описательно: Z_α, Z_β, Z_γ, . . ., можно обозначить короче: $Z_{\alpha\cdot\beta\cdot\gamma}\ldots$, что следует понимать так, что глаголу Z присущи, как лексеме, одновременно и признак α, и признак β, и признак γ и т. д.

1) dürfen	Z_0	2) heissen	$Z_{100\cdot001\cdot0}$
können	Z_0	3) lassen	$Z_{100\cdot010\cdot0}$
mögen	Z_0	4) haben	$Z_{100\cdot1}$
müssen	Z_0	5) sein	$Z_{000\cdot001\cdot1}$
sollen	Z_0	6) werden	$Z_{001\cdot0\cdot1}$
wollen	Z_0		

Теперь мы имеем все необходимое для того, чтобы сформулировать модель.

Соответствия D	Соответствия D	Соответствия İ	Образцы	Примеры	Число случаев
0	0	0	$S\hat{}Л(Z_{000})$	Die Reihe konwergiert	226
		1	$S\hat{}Л(Z_{001})\hat{}P$	Der Teil scheint reel	360
	1	0	$S\hat{}Л(Z_{100})\hat{}O_1$	Beschreibungen schließen den Band	575
		1	$S\hat{}Л(Z_{101})\hat{}O_1\hat{}P$	Der Verfasser setzt den Verlauf (dem Werte) proportional	12
1	0	0	$S\hat{}Л(Z_{010})\hat{}O_2$	Die Funktion genügt einer Bedingung	103
		1	$S\hat{}Л(Z_{011})\hat{}O_2\hat{}P$	Die Kurve geht als die Funktion in das Problem ein	1
	1	0	$S\hat{}Л(Z_{110})\hat{}O_1\hat{}O_2$	F nähert sich einer Asymptoten	53
		1	$S\hat{}Л(Z_{111})\hat{}O_1\hat{}O_2\hat{}P$		0
				Всего:	1330

II. СИСТЕМА ТРАНСФОРМАЦИЙ

1. Пассивная трансформация.

а) Заменить глагол, стоящий в личной форме, глаголом с признаком $Z_{001\cdot 1}$ (т. е. безразлично *werden* или *sein*); глагол, стоявший в личной форме, поставить в конце цепочки в форме причастия второго.

б) Заменить слово, стоящее в форме подлежащего, той же формой слова, стоящего в форме прямого дополнения; слово, стоявшее в форме подлежащего, опустить.

2. Модальная трансформация.

Заменить глагол, стоящий в личной форме, той же формой глагола с признаком Z_0 (т. е. на выбор глаголом *dürfen*, *können*, *mögen*, *müssen*, *sollen*, *wollen*, *heissen* или *lassen*); глагол, стоявший в личной форме, поставить в конце цепочки в форме инфинитива.

3. Перфектная трансформация.

Сделать одно из двух:

а) либо заменить глагол с признаком $Z_{1\text{дiv}\,0}$ (т. е. глаголы с семантическим множителем прямой переходности плюс модальные глаголы), стоящий в личной форме, той же формой глагола с признаком $Z_{100\cdot 1}$ (т. е. *haben*), после чего глагол, стоявший в личной форме, поставить в конце цепочки в форме причастия второго;

б) либо заменить глагол с признаком $Z_{0\text{дi}}$ (т. е. не прямо переходный глагол [2]), стоящий в личной форме, той же формой глагола с признаком $Z_{000\cdot 001\cdot 1}$ (т. е. *sein*), после чего глагол, стоявший в личной форме, поставить в конце цепочки в форме причастия второго.

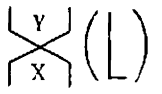
4. Временна́я трансформация.

Заменить глагол, стоящий в личной форме, той же формой глагола с признаком $Z_{001\cdot 0\cdot 1}$ (т. е. *werden*); глагол, стоявший в личной форме, поставить в конце цепочки в форме инфинитива.

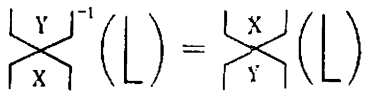
[2] Факт спряжения некоторых непереходных глаголов с *haben* пока, для упрощения, игнорируем.

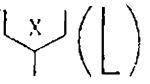
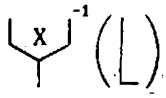
Трансформации пассивную и перфектную можно называть причастными, модальную и временную — инфинитивными.

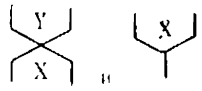
Поскольку трансформации представляют собою в достаточной степени формализованные, хотя и словесные правила, постараемся выразить их символически. Для этого введем два оператора. Оператор замены



означает, что в цепочке L вместо символа X надо записать символ Y. Полагаем:



Оператор записи-стирания  означает, что в цепочке L на некотором месте (в частности — на n+1 месте, где n — число символов цепочки) надо записать символ X. Обратный оператор  означает, что в цепочке L надо стереть некоторый символ X. Используя таким образом два оператора, введенные здесь

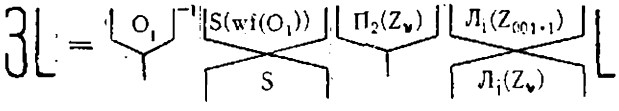


а также три оператора математической логики

$\overline{X}$ — отрицание X,
XvY — сумма X и Y, оператор «или»,
X · Y — произведение X и Y, оператор «и», —

запишем наши 4 правила трансформации, или (в порядке их следования) операторы З, М, В, F [3].

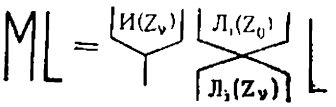
1. Оператор З (пассивная трансформация):



---

[3] Дополнительные обозначения:
1) W — слово (лексема);
2) если $f_j(W_\beta)$ — некоторая форма слова W_β, то wf (f_j) — слово, стоящее в данной форме f_j.
3) i, j, ... — индексы форм; α, β — индексы лексем.
4) Вместо T (L) = N (K(Q(L))) пишем TL = NKQL.

2. Оператор M (модальная трансформация):



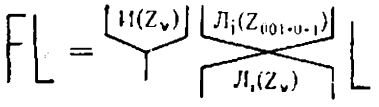
3. Оператор B (перфектная трансформация):

$$B=(B_1 \vee B_2)\cdot\overline{(B_1 \vee B_2)},$$

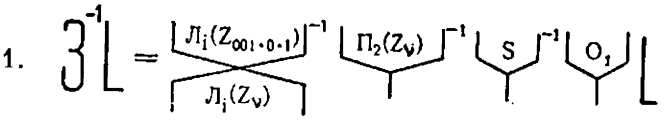
где

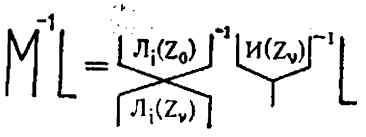


4. Оператор F (временная трансформация):

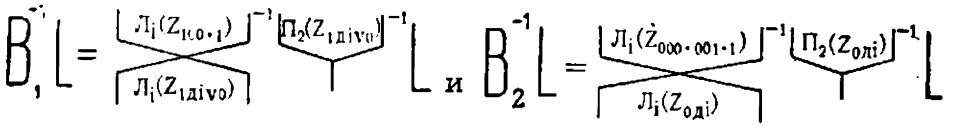


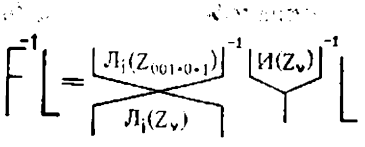
Определим также операторы, обратные данным, что нетрудно сделать сразу в символической форме.

1. 

2. 

3. $B^{-1}=(B_1^{-1} \vee B_2^{-1})\cdot\overline{(B_1^{-1} \vee B_2^{-1})}$, где



4. 

Итак, мы имеем систему аксиом A и систему трансформаций T (и, кроме того, определены трансформации, обратные данным). Прилагая к любой аксиоме любые трансформации в различной последовательности и числе, мы можем получить любой производный образ (синтез). Наоборот, имея некоторый производный образ (а зная образ, мы знаем тем самым историю его синтеза, т. е. последовательности трансформаций, приложенных к некоторой аксиоме для его получения), можно, прилагая к нему в обратном порядке обратные трансформации, получить исходную аксиому (анализ). Поскольку трансформации, по своему определению,

приложимы к любому образу, а не только к элементарному, множество производных образов, синтезируемых грамматикой [A, T], будет бесконечным, если не наложить на применение трансформаций какие-либо ограничения. Наблюдения над текстами позволяют сделать следующие два допущения относительно применимости прямых трансформаций (допущения лишь приблизительные, как и прочие закономерности нашей модели):

1) каждая прямая трансформация может использоваться при образовании данной цепочки из аксиомы только один раз;

2) каждая трансформация может применяться только либо к аксиоме, либо к образу — результату применения трансформации, стоящей левее данной в ряду: З, М, В, F.

Эти ограничения означают практически следующее. Если имеется аксиома, то от нее можно образовать все 4 производных образа первого порядка. От каждого из них можно образовать максимум по 3 производных образа второго порядка (выбывает оператор З). От каждого из них можно образовать максимум по 2 образа третьего порядка и от последних — максимум по одному образу четвертого порядка. Следовательно, если обозначить производный образ — результат применения некоторой трансформации Т (К), или проще ТК, то от одной аксиомы производимы лишь следующие образы:

Нулевого порядка (аксиомы)	Первого порядка	Второго порядка	Третьего порядка	Четвертого порядка
A	ЗА	МЗА	ВМЗА	FВМЗА
			FМЗА	
		ВЗА	FВЗА	
		FЗА		
	МА	ВМА	FВМА	
		FМА		
	ВА	FВА		
	FА			
1	4	6	4	1

Таким образом, каждое подмножество образов, производных от одной аксиомы, насчитывает 16 образов (включая аксиому). При 8 аксиомах это составит 128 образов. Таково максимальное число образов простых предложений немецкого языка, допускаемое данной моделью; эту модель, следовательно, можно представить в виде прямоугольной таблицы из 8×16 клеток. Анализ текста на последнем этапе состоял в том, что множество из 2000 независимых предложений было рассортировано по клеткам этой таблицы и в каждой клетке записывалось число случаев. Такая таблица с результатами подсчета и примерами предложений (препараты!) приводится в приложении II. За недостатком места часть таблицы, содержащая клетки для образов 3-го и 4-го порядков, опущена. Это ничего не меняет, так как примеров на образы таких порядков в обследованном тексте не нашлось и соответствующие клетки остаются пустыми.

Одним из достоинств формального подхода в любой науке является возможность четкой экспериментальной проверки следствий, вытекающих из положений, которые постулируются на основе содержательного рассмотрения исследуемого материала. Формальным следствием данной модели, определяемой набором аксиом, трансформациями и правилами их приложения, является как раз указанная таблица. Цифры же, стоящие в ее клетках, служат средством сопоставления этого следствия с фактами языка. Из рассмотренного числа предложений (около 2000) 1960, или 98%, нашли свое место в таблице.

Таким образом, первая часть критерия Хомского («Грамматика должна производить все грамматичные предложения данного языка») в грубом приближении удовлетворяется. Что касается второй части («Грамматика не должна производить ни одного неграмматичного предложения данного языка»), то, как об этом свидетельствуют пустые клетки таблицы (их 97 из 128, т. е. 76%), модель в этом отношении совершенно неудовлетворительна.

Свести к минимуму количество неграмматичных предложений, производимых моделью, можно было бы, очевидно, путем расширения системы правил, ограничивающих применение трансформаций. Правила этой системы могли бы иметь, например, следующий вид. Если считать процесс, состоящий в последовательном приложении к аксиоме нескольких данных операторов, новым оператором, то радикальное решение вопроса состоит в том, чтобы сопоставить каждому из полученных таким образом операторов область его определения, т. е. совокупность образов (в частности, производных), к которым он приложим. Ясно, однако, что такая система ограничивающих правил была бы бесполезна по своей громоздкости. Приведенные выше правила — результат попытки заменить такую универсальную, но бесполезную систему системой грубой, но простой. Поиски в этом направлении можно продолжать. Но важно помнить, что успех здесь зависит от удачного выбора самой формы таблицы, т. е. от решения вопроса, что отнести к аксиоме, что — к трансформациям. Недостаточно компактное расположение занятых клеток в нашей таблице и, главное, резкие колебания чисел по строкам и столбцам свидетельствуют о том, что в данной работе этот выбор далек от оптимального. Об этом говорит также и то, что в числе 30 предложений, не нашедших места в таблице, большинство составляют модели типа Die Ergebnissen sind zu erklären, с формулой S^Л ($Z_{000\,.\,001.\,1}$) И^{zu}(Z_{100}), весьма характерные для немецкого языка и встречающиеся нередко также и в трансформированном виде, что свидетельствует об их продуктивности. При заполнении таблицы допущена также натяжка в том отношении, что не учитывались приименные дополнения, т. е. предложения типа Der Kreis ist gross, Der Wert ist gleich dem Werte считались одинаково принадлежащими к классу с аксиомой S^Л (Z_{001})^Р.

Если говорить о других недостатках работы, необходимо отметить и то, что модель не учитывает наглядно морфологической омонимии, фактически существующей у модальных глаголов и глагола *werden*. Последние в качестве знаменательных [4] образуют Partizip II по одной модели, а в качестве вспомогательных — по другой. Поэтому, например, в клетке e4 Partizip II от *werden* стоит в форме *geworden*, а в клетке b7 — в форме *worden*, из образов же не видно, что это одна и та же лексема.

Но на данном этапе мы и не надеялись получить совершенную модель. Результат своей работы автор видит скорее в том, что удалось нащупать методику исследования, позволяющую, построив приближенную схему, объективным образом сравнить выводы из нее с экспериментальными фактами.

Идеи трансформационной грамматики весьма плодотворны и позволяют, по-видимому, расширить область исследования, связанную с данной работой, в двух направлениях. Во-первых, представляется возможным строить аналогично описанному образы сложных предложений. Во всяком случае, по имеющимся наблюдениям, придаточные предложения по своим моделям мало отличаются от независимых. Здесь на первый план выступят, вероятно, вопросы порядка слов, следовательно — операторы перестановки. С другой стороны, представляется заманчивым распространить трансформационный анализ «по вертикали», охватив уровень словоизменения, что позволит связать образ непосредственно с буквенной записью предложения. Для некоторых областей немецкой морфологии построить систему морфологических операторов, по-видимому, не очень трудно.

Значение подобной грамматики, если бы удалось ее построить, очевидно. Особенно много она должна дать, во-первых, для машинного перевода, позволяя свести построение независимого алгоритма к обращению модели грамматики в сторону анализа и в сторону синтеза, а во-вторых, для обучения языку, так как благодаря далеко идущей формализации вся система должна умещаться буквально «на пяти страницах».

ПРИЛОЖЕНИЕ 1

СЛОВАРЬ, ПОЯСНЯЮЩИЙ ПРИНЦИПЫ КЛАССИФИКАЦИИ И КОДИРОВАНИЯ ГЛАГОЛОВ

Слово	Символ
abbilden	Z_{010}, Z_{110}
abhängen	Z_{010}
anknüpfen	Z_{010}
annähern	Z_{110}
annehmen	Z_{100}, Z_{101}
anwenden	Z_{100}, Z_{110}
ausdehnen	Z_{110}
ausgehen	Z_{010}
aussehen	Z_{001}
bedeuten	Z_{101}
bedienen	Z_{110}
beginnen	Z_{100}, Z_{010}, Z_{110}
begnügen	Z_{110}
bekommen	Z_{100}
beruhen	Z_{010}
beschäftigen	Z_{110}
beschränken	Z_{110}
bestehen	Z_{000}, Z_{010}
betragen	Z_{001}

[4] Модальные глаголы Z_0 в этой функции в нашем тексте не встречаются.

Слово	Символ
betten	Z_{110}
bezeichnen	Z_{110}
bleiben	Z_{001}
bringen	Z_{010}
danken	Z_{110}
denken	Z_{101}
deuten	Z_{101}
dienen	Z_{011}
differenzieren	Z_{110}
DÜRFEN	Z_0
eignen	Z_{110}
eingehen	Z_{010}
erfolgen	Z_{011}
entfallen	Z_{010}
entsprechen	Z_{010}
entwickeln	Z_{000}, Z_{010}, Z_{110}
ersetzen	Z_{000}, Z_{110}
erstrecken	Z_{110}
erscheinen	Z_{001}
erweisen	Z_{101}
folgen	Z_{010}
fragen	Z_{010}
führen	Z_{010}, Z_{110}
geben	Z_{100}, Z_{110}
gehören	Z_{010}
gelingen	Z_{010}
gelten	Z_{000}, Z_{010}, Z_{001}
genügen	Z_{010}
HABEN	Z_{100}, Z_1
halten	Z_{110}
handeln	Z_{110}, Z_{000}
HEISSEN	Z_{100}, Z_{001}, Z_0
hinausführen	Z_{110}
hineinspielen	Z_{010}
integrieren	Z_{000}, Z_{100}
interessieren	Z_{000}, Z_{110}
KÖNNEN	Z_0
LASSEN	Z_{100}, Z_{010}, Z_0
liegen	Z_{000}
MÖGEN	Z_0
multiplizieren	Z_{000}, Z_{100}, Z_{110}
MÜSSEN	Z_0
scheinen	Z_{001}
schliessen	Z_{110}
schreiben	Z_{101}
SEIN	Z_{000}, Z_{001}, Z_1
setzen	Z_{000}, Z_{110}, Z_{101}
SOLLEN	Z_0
stossen	Z_{010}
substituieren	Z_{000}, Z_{110}
überführen	Z_{110}
übergehen	Z_{010}
unterscheiden	Z_{110}
verbinden	Z_{110}
verdanken	Z_{110}
verfügen	Z_{010}
verschaffen	Z_{110}
verwenden	Z_{110}
verzeichnen	Z_{110}
wächsen	Z_{010}
WERDEN	Z_{001}, Z_0, Z_1
widersprechen	Z_{010}
widerstehen	Z_{010}
WOLLEN	Z_0
zeichnen	Z_{101}, Z_{110}
zuordnen	Z_{110}
zurückfüren	Z_{110}
zusammenfallen	Z_{010}

Производные образы

	Нулевого порядка	Первого порядка				Второго порядка					
	A	ЗА	МА	ВА	FA	МЗА	ВЗА	FЗА	ВМА	FМА	FВА
	1	2	3	4	5	6	7	8	9	10	11
a	$СЛ(Z_{000})$ die Reihe konvergiert 226	$Л(Z_{001\cdot1})П_2(Z_{000})$ wird berechnet 1	$СЛ(Z_0)И(Z_{000})$ die Entwicklung muss konvergieren 17	$СЛ(Z_{000\cdot001\cdot1})П_2(Z_{000})$ 8	$СЛ(Z_{001\cdot0\cdot1})И(Z_{000})$ wir werden vorgehen 4						256
b	$СЛ(Z_{100})O_1$ Beschreibungen schliessen den Band 575	$СЛ(Z_{001\cdot1})П_2(Z_{100})$ die Bestimmung sei besprochen 283	$СЛ(Z_0)O_1И(Z_{100})$ man kann die Funktion bestimmen 103	$СЛ(Z_{100\cdot1})O_1П_2(Z_{100})$ wir haben die Möglichkeit untersucht 35	$СЛ(Z_{001\cdot0\cdot1})O_1И(Z_{100})$ das Buch wird Anklang finden 14	$СЛ(Z_0)П_2(Z_{100})И(Z_{001\cdot1})$ die Teile müssen berechnet werden 53	$СЛ(Z_{000\cdot001\cdot1})П_2(Z_{100})П_2(Z_{001\cdot1})$ die Bedeutung ist erkannt worden 12	$СЛ(Z_{001\cdot0\cdot1})П_2(Z_{100})И(Z_{001\cdot1})$ der Raum würde aufgespannt werden 2	$СЛ(Z_{100\cdot1})O_1И(Z_{100})П_2(Z_0)$ wir haben keine Vermutungen machen können 2	$СЛ(Z_{001\cdot0\cdot1})O_1И(Z_{100})И(Z_0)$ man wird diesen Vektor verbessern können 2	1082
c	$СЛ(Z_{010})O_2$ die Funktion genügt einer Bedingung 103	$Л(Z_{001\cdot1})O_2П_2(Z_{010})$ wird auf die Möglichkeit hingewiesen 7	$СЛ(Z_0)O_2И(Z_{010})$ sie sollen Lipschitzbedingungen genügen 3	$СЛ(Z_{000\cdot001\cdot1})O_2П_2(Z_{010})$ das Verfahren hat zur Entscheidung geführt 4							117
d	$СЛ(Z_{110})O_1O_2$ F nähert sich einer Asymptoten 53	$СЛ(Z_{001\cdot1})O_2П_2(Z_{110})$ das Kapitel ist dem Verhalten gewidmet 26	$СЛ(Z_0)O_1O_2И(Z_{110})$ F kann sich einer Asymptoten nähern 9	$СЛ(Z_{100\cdot1})O_1O_2П_2(Z_{110})$ die Arbeiten haben sich mit den Schwierigkeiten befasst 2		$СЛ(Z_0)O_2П_2(Z_{110})И(Z_{001\cdot1})$ P kann in das Q überführt werden 5					95
e	$СЛ(Z_{001})P$ der Teil scheint reel 360		$СЛ(Z_0)РИ(Z_{001})$ die Strecke kann als Mass dienen 15	$СЛ(Z_{000\cdot001\cdot1})РП_2(Z_{001})$ sind Sätze möglich geworden 6	$СЛ(Z_{001\cdot0\cdot1})РИ_{(001)}$ die Fachwelt wird dankbar sein 3						384
f	$СЛ(Z_{101})O_1P$ der Verfasser setzt den Verlauf proportional 12	$СЛ(Z_{001\cdot1})РП_2(Z_{101})$ die Arbeiten werden zugänglich gemacht 8	$СЛ(Z_0)O_1РИ(Z_{101})$ man kann sie als Verfahren auffassen 2	$СЛ(Z_{110\cdot1})O_1РП_2(Z_{101})$ die Theorien hatten sich ungeeignet entlassen 2	$СЛ(Z_{001\cdot0\cdot1})O_1РИ(Z_{101})$ die Glieder werden sich als wesentlich erweisen 1	$СЛ(Z_0)РП_2(Z_{101})И(Z_{001\cdot1})$ c und c dürfen gleich gesetzt werden 1					26
g	$СЛ(Z_{011})O_2P$ die Kurve geht als die Funktion in das Problem ein 1										1
h	$СЛ(Z_{111})O_1O_2P$										0
	1130	325	150	57	22	59	12	2	2	2	0 1961

ИРЯ
институт русского языка
им. В.В. Виноградова
www.ruslang.ru

ON MODELLING THE STRUCTURE OF THE SIMPLE SENTENCE

Summary

A system for a formal description of a fragment of German grammar is developed. This fragment is the grammar of configurations of so-called «basic» parts of a simple sentence (subject, verb, direct and indirect object, predicative). In the present paper such configurations are called predicative syntagms. Any such configuration can be deduced from some elementary predicative syntagm by means of certain rules called transformations. A finite set of elementary predicative syntagms, a finite set of transformations, and some rules restricting the application of transformations are proposed. This formal grammar is reversible, i. e. in the set of transformations for each transformation, T, there is a reciprocal one, T^{-1}. So any type of predicative syntagm, L^1, deduced from some other type L by the application of some transformation, T_1, can be reduced back to the original type, L, by the application of T^{-1} to L'.

Some experimental data justifying and partly correcting these formal conceptions are given.

Л. Н. Засорина

МОДЕЛЬ ИМЕННОГО СКЛОНЕНИЯ ДЛЯ РУССКОГО ПИСЬМЕННОГО ТЕКСТА

0. 0. Материалом для настоящего сообщения послужила работа по синтезу именных форм русского языка для машинного перевода [1]. Представляется, что полученные результаты могут быть интерпретированы в более широком плане, в частности в связи с проблемой идентификации морфем.

ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

0. 1. Грамматики, алгоритмы, модели грамматик. Новые приложения лингвистики для целей автоматизации перевода и переработки информации расширяют существующие представления о грамматиках языка. К известному противоположению типов грамматик (описательная, историческая, сравнительная, общая) добавляется ряд разновидностей. Выдвинутое Л. В. Щербой различение активной и пассивной грамматики получило дальнейшее подтверждение в принципиальном разграничении двух типов алгоритмов: анализа и синтеза [2].

Алгоритмы, которые строятся для машинного перевода, а также для преобразования сообщений с естественных языков на информационные и т. п., являются грамматиками особого рода. Существующие грамматики для подобных целей недостаточны в силу их эллиптичности, которая оправдана особенностями человеческой мысли. Человек, пользующийся грамматикой, способен одновременно понимать множество фактов, аналогически их анализировать, опираясь на интуицию, опыт, конкретную ситуацию. При алгоритмическом подходе источники строго ограничены: а) оригинальными текстами, рассматриваемыми как последовательности букв; б) таблицами, заранее подготовленными; в) программами для операций над а) и б) [3].

Алгоритмы, так же как и грамматики, целесообразно рассматривать как более или менее совершенные модели языка, приближающиеся в разной степени к полной грамматике. Очевидно, что составить всеобъемлющую грамматику, с учетом всей совокупности бытующих слов, стилей, жанров, в настоящее время едва ли возможно и необходимо. Принципиально воз-

---

[1] Основные положения работы изложены в докладе Л. Н. Засориной и Г. С. Цейтина «Алгоритм синтеза именных форм русского языка», прочитанном на Межвузовской научной конференции по прикладной лингвистике в г. Черновцы, 1960.

[2] В. П. Берков. О пассивной грамматике. «Вестник Ленинградского университета», № 20, Серия истории, языка, литературы, вып. 4. 1960.

[3] A. G. Oettinger. Automatic language translation. «Lexical and technical aspects with particular reference to Russian. Harward Monographs in applied science», № 8. Cambridge, Massachusetts, 1960.

можно построить разные модели грамматик. Они будут различаться прежде всего по охвату языкового материала. Так, в практике машинного перевода признано целесообразным составление алгоритмов для анализа и синтеза текстов либо письменного языка в целом, либо научного текста, либо определенной области науки, например математики, физики, радиотехники, медицины и т. п.

С другой стороны, модели могут быть различными в зависимости от общей установки их составителей, принципов и методов анализа языка, строго формальных или формально-семантических и др. Ср., например, структурные и дескриптивные грамматики [4] и «традиционные», академические грамматики.

Что касается алгоритмов, то выбор схемы для них определяется еще их техническими приложениями, данными электронно-вычислительных машин (объем памяти и быстродействие).

0. 2. Письменная и устная формы языка. Существующие грамматики русского языка не ориентированы строго на анализ или синтез речи, хотя ближе по типу к синтезирующим алгоритмам. Также не предназначаются они для какой-либо области текстов. При написании грамматики не принято как-либо ограничивать сферу языкового материала, как это обычно, например, для словарей (областные, специальные, толковые, краткие и полные и т. д.). Между тем потребность в подобных ограничениях весьма актуальна. Так, размежевание письменной и устной формы общения не слабеет к настоящему времени. Возрастающий прирост научной информации ведет к дальнейшей специализации языка в соответствующих областях науки. Возникают усложненные системы символов, «символические языки», например, в химии, геометрии и т. д. [5] Парадоксально, что в определенную историческую эпоху устная форма общения была единственной, а в современном обществе возможно существование «языков» только в письменной форме, не озвученных, ср. языки математической логики, информационные языки и др.

Естественно, что прогресс цивилизации оказывает мощное воздействие на устную речь, в том числе и разговорную. Но разрыв между письменным и устным языком никогда не может быть ликвидирован. Возникает задача более четкого разграничения между ними, которое, по-видимому, является весьма динамичным. Предварительный анализ показывает, что грамматика письменного русского языка имеет отличительные особенности по сравнению с грамматикой устного языка [6]. В частности, такой грамматике свойственна наибольшая стандартизация, здесь складываются особые парадигмы словоизменения и т. д.

0. 3. Принятые ограничения. Предлагаемая модель именного склонения рассчитана на синтез падежных форм; существительного (С), прилагательного, полного и краткого, порядковых числительных и некоторых местоимений (А).

Пределы применимости модели ограничиваются письменным научным текстом, в частности по математике.

Исходным принципом при построении схемы было предположение, что в естественных языках наличествуют более или менее строгие системы правил, порождающих единицы языка на разных уровнях структуры.

---

[4] L. Hjelmslev. La catégorie des cas, étude de grammaire générale. «Acta Jutlandica», Aarhus, VII, 1935 (1), IX, 1937 (2); H. Hockett. Two Models of Grammatical Description. «Word», 1954, № 2/3.

[5] О креолизации языковых систем см.: В. В. Иванов. Язык в сопоставлении с другими средствами передачи и хранения информации. «Тезисы докладов на конференции по обработке информации, машинному переводу и автоматическому чтению текста». М., 1961.

[6] Т. М. Николаева. Типологическое сопоставление русского устного и письменного языков. «Тези доповідей міжвузівської наукової конференції, 22—28 вересня 1960 р». Черновцы, 1960.

Поэтому модели грамматик могут быть получены в результате внимательного изучения зависимостей и парадигм в самом языке. Такой подход, по-видимому, даст наименее объемные и сложные схемы таблиц и правил.

Объектами рассмотрения являются падежные словоформы как цепочки букв. На основе графического анализа производится расчленение словоформ на сегменты: окончания и основы. Анализ ведется, по возможности, без обращения к звуковому составу морфем. В связи с этим допускается «нетривиальное» выделение морфологических элементов, например, окончания *-ь*, *-й*, в словоформах *корень*, *гений*.

Примем, что образование форм имени производится по запросу, поступающему извне, например из предыдущей части программы синтеза текста, в частности после формирования дерева предложения. Предварительно составляется словарь основ, в запросе указан номер основы и падежная форма, которую требуется сформировать. В этих условиях работа над моделированием грамматики (или раздела грамматики) сводится к построению таблиц окончаний для падежных форм, определению формальных признаков основ, диктующих выбор определенной парадигмы, и определению правил, регулирующих извлечение информации из таблиц и словаря и, если требуется, ее преобразование.

Имя существительное.

1. 0. Материал научных текстов подсказывает возможное упрощение парадигмы существительного за счет винительного падежа. Одушевленные имена здесь почти не встречаются. В предлагаемой схеме вин. пад. исключается как самостоятельная единица во всех случаях, кроме сущ. жен. рода (CII), где выбор окончания в вин. пад. не определяется признаком одушевленности-неодушевленности. По тем же основаниям не учитываются разновидности окончаний в род. и предл. пад. типа: *чая*, *чаю*, *в лесу*, *в лесе*.

1. 1. В русском языке размежеваны парадигмы существительных ед. и множ. числа, в пределах ед. числа — женские и не женские типы, а в последних муж. и ср. род различаются по форме им. пад. Информация о ед. числе по существу избыточна, ее можно заменить указанием признаков рода. В словаре основ имеем всего четыре признака рода—числа: *m* — муж. род, *n* — ср. род, *f* — жен. род, *p* — множ. число.

В грамматиках русского языка выделяется 4 основных парадигмы склонения существительных: первое (I), второе (II), третье (III) склонение и парадигма множ. числа (р). Каждая из них расщепляется на твердую и мягкую разновидность. «Неправильные» парадигмы нами не учитываются, поскольку «особенности» и «исключения» ограничены главным образом лексикой бытового обихода: *соседи*, *черти*, *око*, *ухо*, *чудо*, *небо*, *теленок*, *галчонок*, *мать*, *дочь*, *бремя*, *семя* и некоторые другие.

Твердые и мягкие парадигмы допускают объединение; тогда для большинства падежных форм, кроме им. пад. CI, предл. CI и CII, дат. CII и род. $C_р$, получаем по два окончания (см. табл. 1_c).

Вариация окончаний, как видно, стандартна, используются четыре пары противопоставленных друг другу буквы (гласных): $\frac{а}{я}$, $\frac{о}{е}$, $\frac{ы}{и}$, $\frac{у}{ю}$. В ряде форм противопоставленные гласные являются окончаниями (ср.: *интеграл-а*, *корн-я*, *интеграл-у*, *корн-ю*, *сумм-ы*, *лини-и*, *сумм-у*, *лини-ю*). В других формах они могут рассматриваться как «предокончания» (ср.: *интеграл-о-м*, *корн-е-м*, *сумм-о-й*, *лини-е-й*, *интеграл-а-м*, *корн-я-м*, *сумм-а-ми*, *лини-я-ми* и т. д.

Анализ показывает, что выбор варианта окончания и «предокончания» определяется не морфологическими факторами, а главным образом фонетической сочетаемостью звуков, в частности твердостью или мягкостью основы, ее конечного элемента. Твердость и мягкость основы (признак звукового яруса) не получает прямого отражения в графической развертке слова (на уровне букв). Поэтому для модели, опирающейся на формально-

Таблица 1с

Падеж	CI	CII	CIII	$C_р$
Именительный	$\frac{0}{-ь,\ -й}$ $\frac{-о}{-е}$	$\frac{-а}{-я}$	-ь	$\frac{-ы}{-и}$ $\frac{-а}{-я}$
Родительный	$\frac{-а}{-я}$	$\frac{-ы}{-и}$	-и	$\frac{-о}{-е}$ -в $\frac{0}{-ь,\ -й}$ -ев, -ей
Дательный	$\frac{-у}{-ю}$	-е, -и	-и	$\frac{-а}{-я}$ -м
Винительный		$\frac{-у}{-ю}$		
Творительный	$\frac{-о}{-е}$ -м	$\frac{-о}{-е}$ -й	-ью	$\frac{-а}{-я}$ -ми
Предложный	-е, -и	-е, -и	-и	$\frac{-а}{-я}$ -х

графические особенности слов, необходимо задать в словарь признаки: твердости (ъ) и мягкости (ь) основы.

Наибольшее разнообразие формантов характеризует им. пад. CI (муж. и ср. рода). Здесь обнаруживается еще одно, пятое противоположение графем: $\frac{(нулевое\ окончание)}{ь,\ й}$ например: $\frac{место}{поле} — \frac{о}{е}$, $\frac{интеграл}{корень,\ гений} \frac{0}{ь,\ й}$.

В паре $\frac{0}{ь}$ выбор окончания, как и в остальных парах гласных, регулируется твердостью—мягостью основы. Буквенное окончание -*й* всегда сочетается с основами, оканчивающимися на гласный. Ср.: *герой, гений*. Окончания -*е*, -*и* в дат. пад. CII и предл. пад. CI и CII обладают валентностью на предшествующую букву (конечную букву основы): -*и* сочетается только с предшествующим -*и* — (например: *гении, здании, линии*), с остальными основами сочетается окончание -*е* (*сумме, корне, месте* и т. п).

Таким образом, существуют стандартные закономерности сочетания противопоставленных друг другу гласных окончаний и «предокончаний» с основами. Валентные признаки основ в неявном виде представлены в словаре. Для строгости целесообразно задать эти признаки в явном виде, обозначив их соответстующими знаками: $Ш_1$ — основы с конечным -*и*-, $Ш_2$ — основы с конечным гласным (*о*-, *у*-, *и*-).

1.2. В парадигме множ. числа наиболее многообразны окончания им. пад. При этом пары $\frac{ы}{и}$, $\frac{а}{я}$ противополагаются не по признакам твердости—мягкости или каким-либо иным валентностям самих окончаний. Ср., например,

$$\frac{интегралы,\ суммы}{корни,\ линии},\ но\ \frac{места,\ бока}{поля,\ тополя}$$

Различение рода в этом случае сокращает вариантность; так, существительные жен. рода имеют окончания $\frac{ы}{и}$, а ср. рода преимущественно $\frac{а}{я}$ (единичные существительные ср. рода оканчиваются на -*и*: *зернышки, колечки, плечи, очки, яблоки* и др.). Для существительных муж. рода остаются обе возможности. В современном русском языке растет

тенденция к распространению окончаний $\frac{а}{я}$ на месте прежних $\frac{ы}{и}$ [7]. Поскольку нет четкого правила для выбора окончаний у существительных муж. и ср. рода, необходимо проверять каждую основу на соединимость с той или другой парой гласных и в словаре хранить особый признак ($А$ — признак окончания $\frac{а}{я}$).

Так же нестандартно синтезируются окончания в род. пад. множ. числа. Академическая грамматика различает здесь три типа окончаний: а) нулевое (*задач, мест, земель*); б) *-ов* (*-ев*) (*интегралов, роев, гениев*); в) *-ей* (*корней, сходимостей*)[8]. Для случая а) соответственно принятым принципам добавляем окончание *й-* и *-ь* (*гени-й, лини-й, земел-ь*), которые в грамматике считаются элементами основы, поскольку морфологическое деление проводится на уровне звуков, а не графем, как в данной модели. Для окончаний типа а) используем триаду $\frac{0}{ь,\ й}$, уже введенную в схему.

Правила для различения окончаний, изложенные в Академической грамматике, невозможно перенести в модель по ряду причин: во-первых, в силу нестрогости формулировок: так, правило для случая в) покрывает *большинство форм, но допускает и иные случаи, ср.: «окончание -ей* получают в род. пад. множ. числа имена существительные муж. и жен. рода с основой на мягкие согласные и шипящие» [9], но *земля—земель, шестерня—шестерен* и т. п.; во-вторых, в силу обращения к фонетическим признакам, которые неоднозначно отражаются в орфографии: ср. правило для типа б): «окончание *-ов* (*-ёв*), *-ев* в род. пад. множ. числа имеют имена существительные муж. рода с основой на твердые согласные и на j («йот»). . .» [10]; при этом конечный основы передается либо буквой *й* (*гений, герой*), либо нулем буквы (*гениев, героев*); в-третьих, в силу избыточности проверки некоторых признаков: так, правила для типа а) и б) предусматривают проверку окончаний в им. пад. ед. числа.

В этом запутанном, на первый взгляд, смешении окончаний прослеживаются те же закономерности сочетаемости морфем, что и описанные выше. Предлагается ряд последовательных бинарных проверок для однозначного определения окончания:

1) Проверяется признак рода (задан в словаре основ): если муж. рода, то 2 [11], если жен. рода, то 5, если ср. рода, то $\frac{0}{ь,\ й}$.

2) Проверяется (для сущ. муж. рода) признак твердости — мягкости основы: если мягкость, то 3, если твердость, то 4.

3) Проверяется (для сущ. муж. рода с мягкой основой) конечная буква основы: если гласная, то *-ев* (*гениев, героев*), если согласная, то *-ей* (*корней*).

4) Проверяется (для сущ. муж. рода с твердой основой) конечная буква основы: если шипящая, то *-ей* (*патронташей*), если не шипящая, то $\frac{о}{е}$ *-в* (*интегралов*).

5) Проверяется (для сущ. жен. рода) признак склонения: если СIII, то *-ей*, если СII, то $\frac{0}{ь,\ й}$.

К стандартным правилам различения противопоставленных гласных добавляется дополнительно проверка для существительных жен. рода

---

[7] Т. А. Иванова. Из истории именного склонения, «Вопросы языкознания», 1957, № 6.

[8] Грамматика русского языка. М., Изд-во АН СССР, 1952.

[9] Указ. соч., стр. 176.

[10] Там же.

[11] Номера указывают отсылку к следующей проверке.

конечного гласного в основе. Ср., например, *линий*, *струй* (см. блок 15 в приложении). Кроме всего, вводится проверка признаков суперсегментных: для сущ. типа *пальцев*, *дворцов*. Регулятором предокончаний здесь является словесное ударение: слова с основами на шипящий и *ц* и с неподвижным ударением получают *-е*, а с подвижным ударением *-о*. Иначе проверяется безударность окончания (по форме твор. пад. ед. ч.: *па́льцем*, но *дворцо́м*). Соответствующим основам задается признак Б. Тот же суперсегментный признак оказывается определяющим при выборе предокончания $\frac{е}{о}$ у сущ. жен. рода ед. числа, твор. пад. Например: *матрицей*, *задачей*, но *суммой*. Эти основы имеют признак твердости, а в твор. пад. принимают «мягкое» предокончание.

1. 3. Правила синтезирования форм существительных. Если учесть все описанные признаки основ (6 типов признаков, см. приложение), то можно построить такой перечень последовательных шагов для синтеза [12]:

I. В таблице 1 выбирается информация об окончании по признаку падежа (указан в запросе) и типа склонения (указан в словаре при данной основе). В большинстве случаев информация дается сразу, иногда после одной-двух дополнительных проверок (ср., например, им. пад. ед. и множ. числа).

II. В таблице 2 выбирается вариант окончания — предокончания по признакам твердости—мягкости, в некоторых случаях после дополнительных проверок.

III. Восстанавливается беглый гласный (см. блок. 18) [13]. Дополнительные проверки указаны особыми номерами в Т1 и Т2, которые служат адресами соответствующих блоков, указанных в приложении. В блоках в свою очередь может быть указан отсылочный номер — адрес к следующему блоку. В этом случае проводится последовательный выбор информации вплоть до получения пар противопоставленных гласных (они являются входами в Т2) или окончательных, единичных формантов [14].

Адъективные парадигмы.

2. Склонение имен прилагательных (А) в русском языке отличается большей регулярностью, чем склонение существительных. Во всех формах, кроме им. пад. ед. числа прилагательных муж. и ср. рода, окончания разграничиваются по признакам твердости—мягкости основы (см. таблицу $1_А$). В им. пад. ед. числа форманты могут быть выбраны однозначно по признаку рода, за исключением окончания *-ой* (*основн-ой*, *больш-ой*): если муж. род, то $\frac{-ы}{-и}$ *-й*, если ср. род, то $\frac{-о}{-е}$ *-е*. Для словоформ на *-ой* задается особый признак основы (В) и проводится дополнительная проверка (по блоку 12).

Так же как полные прилагательные, синтезируются формы полных причастий. Основам причастий действительного залога (после синтезирования суффикса причастия) приписывается признак мягкости (*проходящий*, *предполагающий*, *проходивший*, *предполагавший*) или твердости (*пройденный*, *предложенный*, *проходимый*).

По схеме полных прилагательных синтезируются и формы порядковых числительных: *первый*, *второй*, *третий* и т. д., основам которых приписываются соответствующие признаки Ъ, Ь, В, $Б_ь$.

---

[12] Соответствующий алгоритм синтеза составлен в основном Г. С. Цейтиным.

[13] Для слов типа: *корень*, *корня*, *палка*, *палок* и др. в словарь задаются признаки беглого *-е-* ($Б_е$) и беглого *-о-* ($Б_о$). Вводится признак беглого *-ь-* ($Б_ь$) для слов типа: *звено*, *звенья*, *крыло*, *крылья*, а также *третий*, *третьего* (см. ниже).

[14] Схема допускает синтез некоторых «неправильностей» в словоформах: *сапогов*, *путью*, *время*, *бойец*. Дополнив модель признаками особенностей для таких слов, можно синтезировать падежные формы в нормативном виде.

Таблица 1_A

Падеж	$A_{m,\,n}$	A_f	A_p
Именительный	$\frac{-ы}{-и}$ *-й, -ой,* $\frac{-о}{-е}$ *-е*	$\frac{-а}{-я}$ *-я*	$\frac{-ы}{-и}$ *-е*
Родительный	$\frac{-о}{-о}$ *-го*	$\frac{-о}{-е}$ *-й*	$\frac{-ы}{-и}$ *-х*
Дательный	$\frac{-о}{-е}$ *-му*	$\frac{-о}{-е}$ *-й*	$\frac{-ы}{-и}$ *-м*
Винительный		$\frac{-у}{-ю}$ *-ю*	
Творительный	$\frac{-ы}{-и}$ *-м*	$\frac{-о}{-о}$ *-й*	$\frac{-ы}{-и}$ *-ми*
Предложный	$\frac{-о}{-е}$ *-м*	$\frac{-о}{-е}$ *-й*	$\frac{-ы}{-и}$ *-х*

В кратких формах прилагательных, причастий и глаголов прош. вр. все окончания также сводятся к противопоставлениям гласных:

$\frac{0}{ь,\ й}$	$\frac{о}{е}$	$\frac{а}{я}$	$\frac{ы}{и}$
нов	*ново*	*нова*	*новы*
хорош	*хорошо*	*хороша*	*хороши*
коротк	*коротко*	*коротка*	*коротки*
(синь)	*сине*	*синя*	*сини* [15]

Представляется возможным слить окончания кратких прилагательных с окончаниями именительного падежа полных прилагательных. Тогда не потребуется вводить особую таблицу, или строку в таблице 1, для синтеза кратких прилагательных. Примем, что парадигма кратких прилагательных представлена единственным падежом — именительным, а все косвенные падежи являются пустыми. Выбор окончания производится при разграничении прилагательных по признаку «краткости» (К) или его отсутствию (К̄) [16].

Данная схема допускает «неправильные» окончания у прилагательных типа: *хороше*, *свеже*, *горяче* вместо *хорошо*, *свежо*, *горячо* (Признак Б, введенный для существительных и прилагательных, диктует здесь выбор *-о* из $\frac{-о}{-е}$ *-е*.) По-видимому, такие формы исключительно редки в научном тексте, однако при надобности схема может быть усовершенствована.

Местоимения.

3. Как известно, местоимения в русском языке не представляют самостоятельных парадигм. Формообразование в некоторых разрядах место-

---

[15] «Мягкие» разновидности кратких форм представлены единичными словами. Окончание *-й* в триаде $\frac{0}{ь,\ й}$ остается незанятым, пустым. Признак беглого *-о-*, *-е-* задается основам прилагательных так же, как и существительным. Для основ с конечной согласной: *к*, *г*, *х*, *ш*, *щ*, как и у соответствующих существительных, предусматривается дополнительная проверка. См. блок 17 в приложении.

[16] Признак «краткости» поступает из запроса синтезировать определенную форму прилагательного.

имений личных, вопросительных (*сколько*, *кто*, *что*), указательных (*этот*, *тот*, *таков*), а также *его*, *ее*, *их* ввиду наличия супплетивизма и других нерегулярностей можно относить к сфере словаря, а не грамматики. Среди остальных разрядов выделяется группа местоимений, падежные формы которых синтезируются, как у полных прилагательных: *каждый*, *самый*, *всякий*, *некий*, *какой*, *каковой*, *который*, *некоторый*, *такой*, *таковой*.

Притяжательные местоимения: *мой*, *твой*, *свой*, *наш*, *ваш* имеют специфику в образовании форм им. пад., а остальные формы совпадают с адъективной парадигмой. Графические формы им. пад. муж. рода делятся на основу *мо-*, *тво-*, *сво-* и окончание *-й*. Особенность форм им. пад. ед. числа ср. и жен. рода и множ. числа в том, что они имеют не «дублетные» окончания *моее*, *мояя*, *моии*, *нашее*, *нашая*, *нашие*, а «краткие» окончания *мое*, *моя*, *наше*, *наша*, *наши* и т. п. Закономерно было бы рассматривать эти формы как принадлежащие к парадигме кратких прилагательных. Тогда *мо-й*, *тво-й*, *сво-й* займут естественное место основ, принимающих окончание *-й* из триады $\frac{0}{ь,\ й}$, которое у кратких прилагательных оставалось пустым. К этой же группе можно присоединить и местоимение *весь*; его основу представим как *вс-*, по признаку беглости (Б) восстанавливается *-е* перед окончанием *-ь*, которое тоже занимает полупустующее место в триаде $\frac{0}{-ь,-й}$. Таким образом, формоизменение притяжательных местоимений и местоимения *весь* представляет собою объединение парадигм краткого прилагательного (для им. пад.) и полного прилагательного (для косвенных пад.). Местоимения с «особенностями» находят себе место, занимают пустые места в описанной схеме. Этот факт подтверждает надежность предлагаемой системы.

4. 1. Рассмотренная модель склонения не устанавливает каких-либо неизвестных ранее закономерностей в формообразовании имен. Последовательно проведенный анализ сочетаемости морфологических единиц и их элементов в плане графического выражения отличает данную схему от «традиционных» грамматик.

Именное склонение в письменной речи (в ее научной разновидности) представлено 7 парадигмами; 4 субстантивными и 3 адъективными. Для различения парадигм и отдельных падежей в языке используется пять пар противопоставленных друг другу «гласных» ($\frac{-а}{-я}$, $\frac{-о}{-е}$, $\frac{-у}{-ю}$, $\frac{-ы}{-и}$, $\frac{0}{-ь,-й}$) наряду с ограниченным количеством других буквосочетаний (*й*, *ю*, *м*, *х*, *ми*, *му*). Обнаруживаются межпарадигматические закономерности, регулирующие выбор той или иной «гласной» окончания или предокончания. Эти закономерности лежат на уровне буквенных (графических) и звуковых, фонетических валентностей элементарных единиц речи (сочетание основ на твердый согласный с буквенными окончаниями *-а*, *-о*, *-у*, *-ы*, но не *-я*, *-е*, *-ю*, *-и*). Таким образом, различение твердых и мягких парадигм и окончаний не принадлежит морфологическому уровню в собственном смысле (формам склонения), а относится к иерархически подчиненному ему фонетическому и буквенному ярусу языка. В связи с этим закономерно выделить более обобщенные морфологические единицы, чем их принято выделять: окончания $\frac{-а}{-я}$*-м*, $\frac{-а}{-я}$*-ми*, $\frac{-а}{-я}$*-х*, но не *-а*, *-я*, *-ам*, *-ям*, *-ами*, *-ями* и т. д.

При последовательном разграничении уровней: а) буквенно-сегментного, б) фонетического (суперсегментного), в) собственно морфологического можно построить правила для преобразования единиц одного уровня в единицы другого уровня. Соответственно этому предлагается синтези-

ровать падежные формы имени в несколько ступеней: 1) определение морфологической единицы (выбор окончания заданной словоформы); 2) определение фонетических валентностей элементов морфемы; 3) развертка окончания в цепочку букв. При этом третий шаг отражает закономерности письменного текста, второй — устной речи, первый — языка в понимании Ф. де Соссюра, иначе метаязыка, ступени конструктов [17].

Падежные окончания в русском языке имеют двухчленный состав в адъективных парадигмах, в парадигме сущ. множ. числа и твор. пад. ед. числа сущ. В них выделяются конечные буквы (или их цепочки) и «предбуквы». В роли предбукв выступают противопоставленные друг другу гласные: $\frac{а}{я}$, $\frac{о}{е}$, $\frac{у}{ю}$, $\frac{ы}{и}$. Они являются дополнительными дифференциальными признаками целых парадигм и отдельных падежных форм. Так, формы существительных множ. числа противостоят формам прилагательных множ. числа:

$$\begin{matrix} С_р & \longleftrightarrow & А_р \\ \frac{-а}{-я} & \longleftrightarrow & \frac{ы}{и} \end{matrix}$$

(исключение составляют лишь существительные им. пад. и род. пад., которые особо выделены и во всех остальных парадигмах). Изучение синкретизма русских падежей должно строиться с учетом не только окончаний в целом, но и их «предбукв». Можно предполагать, что закон унификации «предбукв» пробивает путь в парадигме существительных множ. числа. По-видимому, этот фактор подкрепляет тенденцию к распространению форм на $\frac{-а}{-я}$ среди существительных муж. рода (ср.: *вектора, обертона, кондуктора, срока* и др.).

ПРИЛОЖЕНИЕ

Таблица 1

Падеж	$С_I$	$С_{II}$	$С_{III}$	$С_р$	$А_{т,\ п}$	$А_f$	$А_р$
Именительный	I	$\frac{а}{я}$	*ь*	2	3	4	5
Родительный	$\frac{а}{я}$	$\frac{ы}{и}$	*и*	6	$\frac{о}{е}$ *го*	$\frac{о}{е}$ *й*	$\frac{ы}{и}$ *х*
Дательный	$\frac{у}{ю}$	7	*м*	$\frac{а}{я}$ *м*	$\frac{о}{е}$ *му*	$\frac{о}{е}$ *й*	$\frac{ы}{и}$ *м*
Винительный		$\frac{у}{ю}$				$\frac{у}{ю}$ *ю*	
Творительный	$\frac{о}{е}$ *м*	$\frac{о}{е}$ *й*	*ью*	$\frac{а}{я}$ *ми*	$\frac{ы}{и}$ *м*	$\frac{о}{е}$ *й*	$\frac{ы}{и}$ *ми*
Предложный	7	7	*и*	$\frac{а}{я}$ *х*	$\frac{о}{е}$ *м*	$\frac{о}{е}$ *й*	$\frac{ы}{и}$ *х*

[17] О понятии конструкта см.: С. К. Шаумян. Преобразование информации в процессе познания и двухступенчатая теория структурной лингвистики. «Тезисы докладов на конференции по обработке информации, машинному переводу и автоматическому чтению текста». М., 1961.

Таблица 2

	ъ	ь
$\frac{0}{ъ,\ й}$ 18	0	15
$\frac{0}{я}$ 19	*а*	*я*
$\frac{у}{ю}$	*у*	*ю*
$\frac{о}{е}$ 19	16	*е*
$\frac{ы}{и}$	17	*и*

Блоки к таблицам 1 и 2

1) *т* — $\frac{0}{ъ,\ й}$ *п* — $\frac{о}{е}$

2) А — $\frac{а}{я}$ $\overline{А}$ — $\frac{ы}{и}$

3) К — 8 $\overline{К}$ — 9

4) К — $\frac{а}{я}$ $\overline{К}$ — $\frac{а}{я}$ *я*

5) К — $\frac{ы}{и}$ $\overline{К}$ — $\frac{ы}{и}$ *е*

6) *т* — 10 *j* — 11 *п* — $\frac{0}{ъ,\ й}$

7) $Ш_I$ — *и* $\overline{Ш}_I$ — *е*

8) *т* — $\frac{0}{ъ,\ й}$ *п* — $\frac{о}{е}$

9) *т* — 12 *п* — $\frac{о}{е}$ *е*

10) Ь — 14 Ъ — 13

11) $С_{III}$ — *ей* $\overline{С}_{III}$ — $\frac{0}{ъ,\ й}$

12) $\overline{В}$ — *ой* В — $\frac{ы}{о}$ *и*

13) $Ш_3$ — *ей* $\overline{Ш}_3$ — $\frac{о}{е}$ *е*

14) $Ш_2$ — *ея* $\overline{Ш}_2$ — *ей*

15) $Ш_2$ — *и* $\overline{Ш}_2$ — ь

16) Б — *е* $\overline{Б}$ — 0

17) $Ш_4$ — *и* $\overline{Ш}_4$ — ы

18) $Б_е$ — *е* $Б_о$ — *о* $Б_ø$ — 0

19) $Б_ь$ — ь $\overline{Б}_ь$ — 0

Признаки основ:

1) $С_I$, $С_{II}$, $С_{III}$, **А**.
2) *т*, *п*, *f*, *р*
3) Ъ, Ь. Признак Ъ присваивается основам существительных с окончаниями и основам прилагательных с предокончаниями: *а*, *е*, *у*, *ы*, 0.

4) $Ш_1$, $Ш_2$, $Ш_3$, $Ш_4$. Перечни конечных букв основы.
$Ш_1$ — основы сущ. на *-и-* (*здание, гений, операция*);
$Ш_2$ — основы сущ. на гласный (*здание, струя, герой*);
$Ш_3$ — основы сущ. муж. рода на *ч, ц, ш, щ, ж* (*нож, патронташ, плащ*);
$Ш_4$ — основы сущ. и **А** на *к, г, х, ч, щ, ш, ж* (*нож, задача, короткий, плохой*).

5) А, Б, В. Устанавливаются по окончаниям в отдельных падежах
А — у сущ. муж., ср. рода и множ. *-а, -я* (*здания, города*); Б — у сущ. и **А** с основой на *ч, ц, ш, щ, ж* и безударным окончанием твор. пад. (*задача, палец, проходивший*);
В — у **А** муж. рода в им. пад. *-ой* (*основной, второй, таковой*).

6) $Б_о$, $Б_е$, $Б_х$. Наличие беглых гласных *е, о* или отсутствие беглости ($Б_х$) перед конечной буквой основы (*корень, палка, короткий, весь*);
$Б_ь$. Наличие «беглого» ь перед окончанием (*звено — звенья, третий — третьи*).

Примеры словарных входов для основ (отмечены только наличествующие признаки):

класс $С_I$mЬ	произвольност $С_{III}$f Ь
операци $С_{II}$f Ь $Ш_1Ш_2$	произвольн **А** Ь
задач $С_{II}$ f Ь Б	втор **А** ЪВ
мест $С_I$n Ь А	*мо(й)* **А** ЬК

A MODEL OF THE NOUN DECLENSION FOR WRITTEN RUSSIAN

Summary

The author attempts to develop a model of the noun declension for Russian. The application of the model is limited to written scientific texts. It can be used for the synthesis of the case forms of nouns, adjectives, ordinals and some of the pronouns. Some tables of endings for the case forms have been compiled and some formal criteria for identifying the type of stem have been defined. The consequent registration of the valency of the morphemic units and their elements allows to reduce the whole system of declension to seven paradigms. Paradigmatic distinctions are brought out by means of five oppositional «vowels»: $\frac{а}{я}$, $\frac{о}{е}$, $\frac{у}{ю}$, $\frac{ы}{и}$, $\frac{0}{ь, й}$ and some other letter combinations. The synthesis of the case forms consists of the following stages: 1) choosing the ending of the given case form from the table, 2) finding the phonetic valency of the morphemic elements, 3) developing the morphemic string into the letter string.

IV. СТРУКТУРНАЯ ЛЕКСИКОЛОГИЯ

Ю. Д. Апресян

О ПОНЯТИЯХ И МЕТОДАХ СТРУКТУРНОЙ ЛЕКСИКОЛОГИИ[1]

(на материале русского глагола)

I. НЕКОТОРЫЕ ВОПРОСЫ ТЕОРИИ

Задачей настоящей статьи является разработка формального аппарата для структурного изучения лексических значений.

1. Два способа изучения значений

Для любого кода, даже самого простого, характерно наличие знаков, т. е. двусторонних единиц, имеющих план выражения (означающее) и план содержания (означаемое). В некоторых кодах, включающих большое число знаков (к ним относятся и все естественные языки), означающее и означаемое обнаруживают известную сложность строения. В этом случае все означающие и все означаемые можно разложить на составляющие их элементарные единицы, которые Л. Ельмслев называет фигурами. В итоге получается относительно ограниченный инвентарь фигур, различные комбинации которых создают реальное многообразие означающих и означаемых.

Фигуры, на которые разлагаются знаки, сами не могут считаться знаками, поскольку они являются односторонними единицами. У фигур плана выражения (фонем) нет плана содержания, а у фигур плана содержания (семантических компонентов, или семантических множителей) нет плана выражения. Фигуры образуют в языке своеобразные вторичные структуры, в которые перекодируется план выражения и план содержания знаков, т. е. фундаментальная структура языка.

Любой из двух планов языка может быть изучен либо на уровне вторичной структуры, т. е. на уровне фигур, либо на уровне первичной структуры, т. е. на уровне знаков.

Это относится, в частности, к единицам плана содержания.

Любое значение может быть рассмотрено как совокупность фигур. Анализ значения на уровне фигур можно проиллюстрировать следующим примером. Пусть у нас имеются слова 1) *отец*, 2) *мать*, 3) *сын*, 4) *дочь*, 5) *дядя*, 6) *тетя*, 7) *племянник*, 8) *племянница*. Сопоставляя 1, 3, 5 и 7 с 2, 4, 6 и 8, получаем фигуру «пол» (а — мужской, ā — женский). Сопоставляя 1, 2, 3 и 4 с 5, 6, 7 и 8, получаем фигуру

[1] Автор пользуется случаем, чтобы выразить искреннюю признательность С. К. Шаумяну, который поделился с ним своими соображениями по существу затрагиваемых в статье вопросов; И. И. Ревзину, с которым автор имел возможность обсудить ряд выдвигаемых в статье положений; К. И. Бабицкому, критические замечания которого оказали автору большую помощь.

«генетическая связь» (b — прямая, b̄ — непрямая). Сопоставляя, наконец, 1, 2, 5 и 6 с 3, 4, 7 и 8, получаем фигуру «поколение» (c — предшествующее, c̄ — последующее). Теперь значение любого элемента в пределах фиксированного нами объекта, т. е. восьми терминов родства, может быть представлено в виде набора из трех символов; ср. *отец* — abc, *мать* — ābc, *сын* — abc̄, *дочь* — ābc̄, *дядя* — ab̄c, *тетя* — āb̄c, *племянник* — ab̄c̄, *племянница* — āb̄c̄.

Как показывает этот пример, анализ значения на уровне фигур позволяет представить значение каждого слова как сложную единицу плана содержания, разложимую на составляющие ее элементарные семантические компоненты.

Анализ значения на уровне знаков не дает информации о составляющих данное значение семантических компонентах. Если, например, мы анализируем те же восемь терминов родства, мы устанавливаем, что все они являются разными знаками, т. е. разными парами означающих и означаемых. На этом основании мы должны рассматривать значение каждого слова (т. е. каждое означаемое) как простую, неделимую на данном уровне единицу плана содержания.

Каждый из двух способов изучения значений имеет определенные достоинства.

Исследование значений на уровне фигур ведет к установлению небольшого числа семантических компонентов, которые входят в большое число конкретных лексических значений. Поэтому описание значений в терминах семантических компонентов может быть в идеале весьма простым (экономным). Однако адекватное разложение означаемых на семантические компоненты возможно далеко не всегда. Оно осуществимо лишь в тех случаях, когда 1) мы имеем дело с однозначными единицами, 2) эти единицы образуют более или менее замкнутую семантическую группу.

Как будет показано ниже, методика анализа значения как элемента знака не ограничена условиями такого рода. Однако описание значений как простых, далее неделимых единиц плана содержания получается гораздо более громоздким, чем описание значений как сложных единиц, разложимых на составляющие их семантические компоненты.

В связи с этим полезно выяснить, имеется ли какая-либо основа для компонентного анализа значений в рамках второй методики (когда значение рассматривается как элемент знака). Очевидно, такая основа будет создана, если мы сумеем не только разграничить значения, но и объединить их по каким-либо структурным признакам в классы эквивалентных значений. Поскольку каждый класс эквивалентных значений должен обладать некоторым инвариантным семантическим свойством, это свойство, очевидно, будет характеризовать и любой член класса. Следовательно, данное свойство может быть приписано любому члену класса в качестве его семантического компонента. Если некоторое значение входит в несколько классов эквивалентных значений, оно может быть разложено на несколько семантических составляющих. Таким образом, мы получаем возможность проводить компонентный анализ значений, хотя имеем дело исключительно со знаками, а не с фигурами. Начав с конкретного лексического значения, мы вновь возвращаемся к нему, но уже на новой основе: теперь мы в состоянии дать достаточно экономное описание многих (однако не всех; см. часть IV) лексических значений.

В свете сказанного конкретная задача настоящей работы сводится к тому, чтобы наметить возможно более формальную методику 1) разграничения лексических значений, 2) объединения лексических значений в классы эквивалентностей, 3) разложения конкретных лексических значений на составляющие их элементарные семантические компоненты. Последний вопрос в данной работе освещается более кратко, чем два первых.

2. Структура и значение

Для выполнения поставленных задач используется аппарат структурной лингвистики, особенно методы дистрибутивного и трансформационного анализа. С помощью этих методов выявляются все различные элементы структуры языка.

Мы исходим из предположения, что различие между любыми двумя лексическими значениями, если оно релевантно в данном языке, отражается в существенных структурных различиях. Таким образом, всякий раз, когда мы устанавливаем существенное структурное противопоставление, мы можем предполагать, что тем самым устанавливается и важное семантическое различие, т. е. различие по крайней мере двух самостоятельных значений [2].

Предлагаемая методика позволяет устанавливать лишь различия и тождества значений, т. е. так называемые дифференциальные или структурные значения. Иными словами, наша методика позволяет строить значимости (в соссюровском смысле слова), но не дает возможности объективно исследовать так называемое сигнификативное и денотативное значение слова (содержание и объем выражаемого им понятия). При этом множество дифференциальных значений может быть не изоморфно множеству сигнификативных и множеству денотативных значений.

Поскольку наша методика создает основу для операционного установления значимых противопоставлений в области лексики, она предназначается для решения некоторых вопросов структурной лексикологии. Она не решает собственно семантических вопросов.

3. Основные допущения

Большинство существующих теоретических моделей, разработанных в структурной лингвистике, построено на предположении, что мы можем каким-то образом отличать отмеченные цепочки символов от неотмеченных. На этом же предположении основан и структурный анализ значений.

(а) О понятии отмеченной фразы.

Понятие «отмеченности» является пока чисто интуитивным, хотя оно, видимо, может быть формализовано.

В современной лингвистике существуют два основных содержательных толкования категории отмеченности. Большинство исследователей, в том числе и Н. Хомский, понимают под отмеченной такую фразу, которая формально правильно построена, но может быть лишена разумного смысла. С этой точки зрения фразы *Бесцветные зеленые идеи бешено спят* и *Я обедаю о тебе* должны быть признаны отмеченными. З. Харрис развил более узкое понимание отмеченности, в силу которого допустимая фраза должна быть не только формально правильно построена («грамматична»), но должна также реализовать типичную для данного языка лексическую сочетаемость. С этой точки зрения фразы *Бесцветные зеленые идеи бешено спят* и *Я обедаю о тебе* должны быть признаны недопустимыми.

---

[2] Сходные допущения содержатся в работах З. Харриса, Н. Хомского и других исследователей, считавших возможным использовать структурные методы для изучения значений. См.: Z. S. Harris. Distributional Structure. «Word», vol. X, № 2—3, 1954; N. Chomsky. Syntactic Structures. 's-Gravenhage, 1957; N. Chomsky. Logical Structure in Language. American Documentation, vol. VIII, № 4, 1957; Z. S. Harris. Co-occurrence and Transformation in Linguistic Structure. «Language», vol. XXXIII № 3, 1957; D. S. Worth. Transform Analysis of Russian Instrumental Constructions. «Word», vol. XIV, № 2—3, 1958; R. Lees. The Grammar of English Nominalizations. «IJAL», 1960.

Эти два понимания отмеченности фразы, внешне как будто несовместимые, фактически не только вполне совместимы, но и дополняют друг друга. В сущности, мы имеем дело не с двумя противоположными точками зрения, а с различной интерпретацией одного и того же понятия на разных уровнях анализа. У Н. Хомского это понятие интерпретировано для чисто грамматического уровня анализа, а у З. Харриса — для чисто лексического уровня. Мы интерпретируем понятие отмеченности для лексико-грамматического уровня анализа языка.

Содержательно отмеченную фразу вообще можно описать как такую фразу, в которой соблюдены предписываемые данным уровнем анализа ограничения. Наиболее абстрактному (высшему) уровню анализа соответствуют наименее сильные ограничения, обусловленные чисто грамматической сочетаемостью морфем. Наименее абстрактному (низшему) уровню анализа соответствуют наиболее сильные ограничения, обусловленные чисто лексической сочетаемостью морфем. Среднему уровню анализа соответствуют ограничения, обусловленные лексико-грамматической сочетаемостью морфем, более сильные, чем на высшем уровне, но менее сильные, чем на низшем уровне.

Необходимо, следовательно, уточнить понятие отмеченной фразы для лексико-грамматического уровня анализа языка. Это понятие должно быть менее широким, чем у Н. Хомского (в частности, из числа отмеченных должны быть исключены фразы типа *я обедаю о тебе*, *он ввергает*, *хороший он*, *радый учитель*, *две мили идутся им*, *я буду написать письмо* и т. п.), но более широким, чем у З. Харриса (в частности, фразы типа *изба идет*, *зеленые идеи*, *любить во все лопатки* и т. п. должны считаться отмеченными).

Мы вынуждены ограничиться этими замечаниями. Строгое определение отмеченной фразы (в лексико-грамматическом смысле) может быть дано лишь после того, как будет построено лингвистическое понятие отмеченности.

б) Установление структуры отмеченной фразы.

Структурный анализ значений может быть проделан после фиксации формальной грамматической модели данного языка. Наличие такой модели позволяет установить структуру отмеченной фразы. Для того чтобы определить структуру отмеченной фразы, необходимо:

1) знать грамматические классы эквивалентных элементов,

2) уметь относить каждое слово фразы к тому или иному грамматическому классу эквивалентных элементов,

3) уметь устанавливать синтаксические связи между словами.

Примем следующее допущение о грамматических классах эквивалентных элементов: допустим, что нам известны функциональные классы слов в терминах модели Ч. Фриза. Это значит, что помимо грамматических классов N, V, A, D [3] нам даны с п и с к о м все служебные элементы (как слова, так и морфемы).

Сделанное допущение является вполне разумным, поскольку такие классы фактически получены для некоторых языков и выработана методика, на основе которой они могут быть получены в принципе для любого языка.

Мы также допускаем, что у нас есть методика, позволяющая нам опознать фразеологические единицы и исключить их из анализа слов.

[3] Мы заменили цифровые символы Ч. Фриза более удобными буквенными символами З. Харриса (см. сноску выше), но не меняли ничего существенного в самой модели Ч. Фриза. Отметим, что предложенное им деление на функциональные классы слов не совпадает с традиционной классификацией слов по частям речи. В N входят все существительные и некоторые разряды местоимений; в V входят глаголы, за исключением вспомогательных, служебных, модальных и связки; в A — все прилагательные, некоторые местоимения и числительные; в D — наречия и некоторые именные группы и т. д.

На основе сделанных допущений можно построить систему структурной лексикологии данного языка [4], пользуясь методами структурной лингвистики, в частности, дистрибутивным и трансформационным анализом.

4. Основные понятия

Вслед за А. И. Смирницким, будем рассматривать слово как множество словоформ, с одной стороны, и множество лексико-фразеологических вариантов, с другой [5].

В дальнейшем мы будем иметь дело с понятием лексико-фразеологического варианта, под которым А. И. Смирницкий понимает слово, рассматриваемое в одном из имеющихся у него значений. Однако термин А. И. Смирницкого является слишком громоздким; поэтому для краткости мы будем говорить не о лексико-фразеологических вариантах слова, а о его лексемах [6].

Формулировку, приведенную в предыдущем абзаце, нельзя, конечно, считать сколько-нибудь строгим определением. Однако для удобства рассуждений мы вынуждены ввести термин «лексема» до того, как описана формальная методика построения «лексем». В предварительном порядке мы можем лишь сказать, что под лексемой понимается такой элемент множества «слово», который выделяется в нем после проведения описываемых ниже операций.

Лексемы и классы эквивалентных лексем (семантические поля) разграничиваются на основе дистрибутивных и трансформационных критериев.

а) Дистрибутивная методика.

Чтобы выразить дистрибуцию лексемы или класса эквивалентных лексем, нам необходимо ввести понятие ориентированной фразы [7], дистрибутивной формулы и конструкции.

Начнем с термина «ориентированная фраза». Поскольку в силу сделанных нами допущений мы можем установить в любой отмеченной фразе все синтаксические связи слов, мы можем, очевидно, построить некоторую новую фразу, ориентированную относительно одного интересующего нас слова (в принципе это может быть любой член N, V, A или D в любом предложении). Если это слово назовем ядром, то под ориентированной фразой будем понимать конечную цепочку слов, каждое из которых синтаксически непосредственно связано с ядром (из совокупности однородных связей удерживается только одна).

Термин «ориентированная фраза» в таком понимании отличается от термина «предложение». В конкретном предложении может быть вскрыто несколько ориентированных фраз в зависимости от того, какое слово задается нам в качестве ядра. В предложении *Он сделался зеленым от страха* выделяется фраза *он сделался зеленым*, ориентированная относительно ядра *сделаться* и фраза *сделаться зеленым от страха*,

---

[4] Соответствующую модель целесообразно строить на основе фиксированного словаря объемом в 5—8 тыс. наиболее частых слов. Такой словарь обеспечивает достаточно полное покрытие лексики любого текста (известно, что 4000 наиболее частых слов покрывают 97,5% любого текста) и в то же время содержит минимум случайных элементов.

[5] А. И. Смирницкий. К вопросу о слове (проблема «тождества слова»). «Труды Ин-та языкознания АН СССР», т. IV. М., 1954.

[6] Мы ввели термин «лексема» в этом необычном для него значении только потому, что не нашли лучшего термина. (Мы не сочли возможным употреблять в данном случае термин «значение», поскольку нехорошо говорить о «дистрибуции» и «трансформации» значения). Можно надеяться, что рано или поздно будет создан метаязык лингвистики, обходящийся без термина «слово». Тогда отпадет надобность и в термине «лексема».

[7] В дальнейшем для краткости мы будем иногда говорить о «фразах», имея в виду ориентированные фразы.

ориентированная относительно ядра *зеленый*. Предложение теоретически бесконечно, а фраза всегда конечна.

Термин «ориентированная фраза» отличается и от термина «словосочетание», поскольку фразой, по определению, должно считаться и предложение.

Выше мы допустили, что умеем относить каждое слово отмеченной фразы к тому или иному грамматическому классу эквивалентных элементов (*в смысле модели* Ч. Фриза). На основании этого допущения и только что введенного понятия ориентированной фразы мы можем определить термин «дистрибутивная формула». Под дистрибутивной формулой понимается ориентированная фраза, все элементы которой, за исключением ядра, представлены на уровне классов слов (т. е. символами, обозначающими тот или иной функциональный класс). Ядро в дистрибутивной формуле представлено конкретным словом (точнее, корневой морфемой). Так, фразе *сделаться зеленым от страха*, ориентированной относительно ядра *зеленым*, соответствует дистрибутивная формула: $\overline{V}$ *зеленым от* N_g [8].

Дистрибутивная формула считается отдельным элементом структуры языка, которому, в силу сделанных нами допущений, соответствует особое значение. В конечном счете каждой лексеме должна быть приписана своя дистрибутивная формула. Поэтому мы можем отождествлять и разграничивать лексемы по их дистрибутивным формулам. Идентичность дистрибутивных формул указывает на тождество лексем, занимающих в них ядерную позицию, а различие дистрибутивных формул является признаком различия лексем, занимающих в них ядерную позицию. Мы получаем, таким образом, возможность устанавливать семантические тождества и различия, т. е. значимости в соссюровском смысле слова.

Под конструкцией понимается ориентированная фраза, все элементы которой представлены на уровне классов слов (т. е. символически). Конструкция, соответствующая фразе *сделаться зеленым от страха*, записывается как $\overline{V}A_i$ *от* N_g.

Конструкция также считается отдельным элементом структуры языка, которому соответствует некоторое обобщенное значение. Каждая конструкция определяет один класс эквивалентных лексем (все те лексемы, которые могут замещать в ней ядерную позицию). Поэтому конструкции могут служить средством разграничения и отождествления классов эквивалентных лексем. Способность всех лексем некоторого набора замещать ядерную позицию в одной и той же конструкции указывает на их принадлежность к одному и тому же классу эквивалентных лексем; если не все лексемы некоторого набора способны замещать ядерную пози-

---

[8] В данной работе применяется обычная нотация, принятая в литературе по дистрибутивному и трансформационному анализу синтаксиса. Значение символов N, V, A и D было уже разъяснено (см. сноску выше). Надстрочные цифры употребляются для разграничения разных членов одного и того же функционального класса. Подстрочные маленькие латинские буквы обозначают падежи. $\overline{V}$ обозначает инфинитив, $V_в$ — безличную форму глагола. Глагольные приставки обозначаются буквой p, ср. pV. $\overset{N}{_{Vn}}$ значит, что данный N получен из V путем междуклассной трансформации (словообразовательного суффикса). S — символ предложения. Предлоги, некоторые окончания и другие структурные элементы удерживаются в формулах в своем нормальном виде.

В связи с символической записью дистрибутивных формул и конструкций необходимо обратить внимание на следующее обстоятельство: символы A, V и другие, как правило, представляют в дистрибутивных формулах и конструкциях не любые, но лишь некоторые члены функциональных классов A, V и т. д. Так, в дистрибутивной формуле $\overline{V}$ *зеленым от* N_g позицию $\overline{V}$ могут замещать глаголы *стать*, *казаться*, *делаться*, но не *читать*, *совершать*, *разрушать* и пр. Объясняется это тем, что на данном уровне анализа учитываются сильные (лексико-грамматические) ограничения, не позволяющие считать фразы типа *разрушить зеленым от страха* отмеченными. Правильно было бы ввести более дробные символы; однако это может быть сделано лишь после завершения анализа в объеме всего фиксированного словаря.

цию в одной и той же конструкции, то это указывает на их принадлежность к разным классам эквивалентных лексем. И в этот случае мы можем устанавливать семантические тождества и различия, т. е. значимости в соссюровском смысле слова.

Все конструкции должны иметь следующий минимальный смысл: если исходное множество содержит n лексем, то максимальное число лексем, способных замещать ядерную позицию в некоторой конструкции, не должно превышать n—1, а минимальное число лексем, способных замещать ядерную позицию в некоторой конструкции, не должно быть меньше 1. В противном случае мы не получим никакого рационального разбиения исследуемого множества элементов (разбиение исходного множества на два подмножества, одно из которых является пустым, в рамках нашей модели представляется нецелесообразным).

Фраза, дистрибутивная формула или конструкция могут быть представлены как цепочки позиций, замещаемых той или иной морфемой или символом. Поскольку любая такая цепочка оказывается ориентированной относительно ядра, мы можем различать во фразах, дистрибутивных формулах и конструкциях три типа позиций: 1) Обязательные позиции, которые при данном ядре должны быть замещены всегда; попытка опустить обязательную позицию приводит к нарушению отмеченности фразы. Примером обязательных позиций могут служить все позиции после слова *ввергает* во фразе *он ввергает меня в пемницу*. 2) Минимые позиции, которые ни при каких обстоятельствах не могут быть замещены по той причине, что их в данной цепочке нет. Примером минимой позиции может служить «позиция» А перед местоимением, ср. *он пишет*, *я читаю* и т. д. 3) От обязательных и минимых позиций принципиально отличаются факультативные позиции, которые могут быть или замещены, или опущены без нарушения отмеченности фразы. Так, во фразе *человек идет* имеются, в частности, следующие факультативные позиции относительно ядра *идет* (они обозначаются круглыми скобками): *человек идет (в столовую) (обедать)*.

Факультативная позиция, которая в данной фразе, дистрибутивной формуле или конструкции опущена, называется свернутой. Факультативная позиция, которая в данной фразе, дистрибутивной формуле или конструкции замещена, называется развернутой. Фраза, дистрибутивная формула или конструкция, в которой, помимо обязательных, есть только свернутые позиции, называется свернутой. Фраза, дистрибутивная формула или конструкция, к которой, помимо обязательных, есть только развернутые позиции, называется развернутой. Фраза, дистрибутивная формула или конструкция, в которой, помимо обязательных есть и свернутые, и развернутые позиции, называется частично развернутой [9].

Возникает вопрос, какую из трех возможных дистрибутивных формул или конструкций (свернутую, частично развернутую или развернутую) мы должны считать диагностической для лексемы или класса эквивалентных лексем. Этот вопрос решается во второй и третьей частях настоящей работы. Здесь важно подчеркнуть другое: наличие в ориентированной фразе, дистрибутивной формуле или конструкции обязательных

[9] Метод свертывания и развертывания конструкций рассматривается Д. Уорфом в его работе о русских словосочетаниях с творительным падежом (см. сноску выше) как равновидность трансформационного анализа. Д. Уорф пользуется понятием нулевой индивидуальной трансформации, которая имеет следующий вид: форма $\rightarrow \sigma$ (прямая трансформация) и $\sigma \rightarrow$ форма (обратная трансформация). Д. Уорф, следовательно, вводит новое понимание термина «трансформация», отличное и от того, которое мы находим в работах Н. Хомского, и от того, которое имеется в работах З. Харриса. Мы не видим оснований для такого нововведения, и в соответствии с общепринятой терминологией относим операции свертывания и развертывания к дистрибутивным. По существу, они являются органической частью анализа по непосредственно составляющим.

и факультативных позиций позволяет представить дистрибуцию лексемы или класса эквивалентных лексем в виде операций свертывания и развертывания [10].

б) Трансформационная методика.

Под трансформацией целесообразно понимать такое преобразование фразы, дистрибутивной формулы или конструкции, в результате которого в ней меняется число или состав только структурных элементов. Это определение является слишком широким, поскольку оно не дает формальных оснований для того, чтобы не считать фразу типа *Волк убивается охотником* трансформом фразы *Волк убивает охотника*.

Однако это определение представляется достаточно узким для того, чтобы отграничивать операции, которые мы будем считать трансформационными, от операций, которые были рассмотрены как дистрибутивные.

В более точном определении трансформации нет пока особой нужды, поскольку трансформации задаются списком. В частности, в данной работе используются следующие трансформации: 1) пассивная (ср. *мы строим дом → дом строится нами*), 2) безличная, ср. *волны швыряют лодку → лодку швыряет волнами*); 3) предложная (ср. *ящик содержит пять предметов — в ящике содержится пять предметов*), 4) видовая (ср. *я напишу письмо → я буду писать письмо*), 5) междуклассная (ср. *я различаю два термина → мое различение двух терминов*), 6) объектно-обстоятельственная (ср. *мы перешли поле → мы шли полем*).

Введем понятие трансформационного комплекса ориентированных фраз, дистрибутивных формул и конструкций.

Пусть у нас имеются фразы $Ф_1$ и $Ф_2$, которым отвечает одна и та же дистрибутивная формула Д. Применим трансформацию Т к каждой из двух фраз. Если их трансформы $тФ_1$ и $тФ_2$ не утрачивают отмеченности, то $Ф_1 \rightarrow тФ_1$ и $Ф_2 \rightarrow тФ_2$ образуют единый трансформационный комплекс ориентированных фраз. В таком случае будем говорить, что Д и ее трансформ тД образуют единый трансформационный комплекс дистрибутивных формул [11]. Если же из двух фраз $Ф_1$ и $Ф_2$ лишь $Ф_1$ сохраняет отмеченность под действием Т, то $Ф_1 \rightarrow тФ_1$ — это один трансформационный комплекс, а $Ф_2 \rightarrow тФ_2$ — это другой трансформационный комплекс. В таком случае будем говорить, что Д и тД не образуют единого трансформационного комплекса.

Пусть, кроме того, у нас имеются фразы $Ф_1$ и $Ф_3$, которым отвечает одна и та же конструкция К. Применим трансформацию Т к каждой из двух фраз. Если они образуют единый трансформационный комплекс ориентированных фраз, то будем говорить, что К и ее трансформ тК образуют единый трансформационный комплекс конструкций. Если же они не образуют единого трансформационного комплекса ориентированных фраз, будем говорить, что К и тК не образуют единого трансформационного комплекса конструкций.

Трансформационный комплекс дистрибутивных формул или конструкций рассматривается как самостоятельный элемент структуры языка. В соответствии с этим основной принцип использования трансформационного анализа для разграничения лексем и классов эквивалентных лексем

---

[10] Операции можно было бы «перевести» в отношения, поскольку между операционной и «реляционной» моделью языка существует отношение изоморфизма (см. Charles F. Hockett. Two Models of Grammatical Description. «Readings in Linguistics» edited by Martin Joos. Washington, 1957, стр. 397). Однако в данном случае удобнее говорить об операциях.

[11] Считается, что дистрибутивные формулы и конструкции не утрачивают отмеченности под действием некоторой трансформации, если им отвечает по крайней мере одна фраза, сохраняющая отмеченность под действием той же трансформации.

можно сформулировать следующим образом: если две или более фразы, которым отвечает одна дистрибутивная формула или одна конструкция, образуют единый трансформационный комплекс, это является признаком тождества лексемы или класса лексем, которые занимают в них ядерную позицию. Если две или более фразы, которым отвечает одна дистрибутивная формула или одна конструкция, не образуют единого трансформационного комплекса, это является признаком различия лексем или классов лексем, которые занимают в них ядерную позицию. Подобно дистрибутивному анализу, трансформационный анализ дает нам возможность устанавливать семантические тождества и различия, т. е. значимости в соссюровском смысле слова.

5. Уровни анализа[12]

Назовем дистрибутивный и трансформационный анализ значений уровнями анализа и выясним, как они соотносятся.

Всего имеется два основных уровня анализа, на которых исследуется значение: морфологический и синтаксический. На синтаксическом уровне выделяются два новых уровня — дистрибутивный и трансформационный. Для удобства можно говорить о морфологическом, дистрибутивном и трансформационном уровне анализа значений. Каждый последующий уровень считается более высоким, чем предыдущий.

Допускается следующее отношение между анализами на разных уровнях: всякое семантическое различие, релевантное в данном языке, может быть описано на высшем уровне анализа [13]; некоторые семантические различия, релевантные в данном языке, могут быть описаны на низших уровнях анализа. Из этого следует, что 1) разграничение единиц на низшем уровне анализа предполагает возможность их разграничения на любом более высоком уровне и поэтому является достаточным основанием для признания установленного различия релевантным в данном языке: 2) тождество единиц на низшем уровне анализа не предполагает с необходимостью их тождества на высшем уровне и поэтому не является достаточным основанием для их отождествления в данном языке; 3) тождество единиц на высшем уровне анализа (т. е. в условиях, когда анализ исчерпан) является достаточным основанием для их отождествления в данном языке.

Следовательно, для установления различий между лексемами или классами эквивалентных лексем достаточно установить их различие хотя бы на одном уровне, а для установления тождества лексем или классов эквивалентных лексем необходим анализ на высшем уровне. Признание двух единиц различными на высшем уровне анализа автоматически уничтожает вывод об их тождестве, который мог быть сделан на более низком уровне.

На первый взгляд, высказанные нами соображения могут показаться противоречивыми. С одной стороны, мы утверждаем, что между элементами структуры (например, морфемами, дистрибутивными формулами, трансформационными комплексами дистрибутивных формул) и значениями

---

[12] Термин «уровень» употребляется в современной лингвистической литературе для обозначения 1) различных ступеней абстракции («чисто реляционный уровень», «субстанциально-реляционный уровень», «субстанциальный уровень»); 2) систем знаков и незнаков (морфологический и синтаксический уровень, с одной стороны, и фонологический уровень, с другой); 3) различных знаковых структур внутри системы знаков (уровень непосредственно составляющих, трансформационный уровень и пр.). Совершенно очевидно, что употребление этого термина должно быть упорядочено. Однако эта важная задача выходит за рамки нашей работы.

[13] Мы предполагаем показать в специальной работе, что даже такие тонкие семантические различия между морфемами, как синонимические различия, могут быть описаны методами дистрибутивного и трансформационного анализа.

можно установить взаимно-однозначное соответствие. С другой стороны, мы говорим, что две единицы, семантически отождествленные нами на одном уровне анализа, могут быть признаны различными на более высоком уровне анализа. В действительности, однако, здесь нет противоречия, так как исключающие друг друга утверждения относятся не к одному, а к разным уровням анализа.

Положение о взаимно-однозначном соответствии элементов структуры и значений имеет обязательную силу лишь для элементов структуры и значений одного уровня, но теряет обязательную силу при сопоставлении элемента структуры одного уровня (например, морфемы) со значением, полученным на другом уровне (дистрибутивном или трансформационном). Иными словами, морфологической единице, дистрибутивной формуле, трансформационному комплексу дистрибутивных формул, конструкции и трансформационному комплексу конструкций всегда можно поставить в соответствие некоторое значение (сему). Однако сема, соответствующая, например, морфеме (т. е. семантический инвариант всех контекстов, в которых данная морфема употребляется), может обнаружить сложность строения на дистрибутивном или трансформационном уровне, т. е. расщепиться на несколько значений. Тогда при переходе с уровня на уровень происходит перекодировка сем.

В настоящей работе значение исследуется только на дистрибутивном и трансформационном уровне.

II. РАЗГРАНИЧЕНИЕ ЛЕКСЕМ

Здесь будут описаны дистрибутивная и трансформационная методика разграничения лексем.

1. Дистрибутивная методика разграничения лексем

Возможны две методики разграничения лексем на основе дистрибутивных критериев. Во-первых, любую дистрибутивную формулу (свернутую, частично развернутую) можно рассматривать как самостоятельный элемент структуры, противопоставленный всем другим элементам. Тогда каждой дистрибутивной формуле необходимо будет приписать самостоятельное значение, отличное от значения любой другой формулы [14]. Во-вторых, можно считать, что только развернутая дистрибутивная формула является самостоятельным инвариантным элементом структуры, поскольку в любом ее варианте (свернутом или частично развернутом) потенциально присутствуют все те позиции, которые фактически замещены в развернутой дистрибутивной формуле.

Рассмотрим подробнее каждую из этих методик.

а) Первая методика.

Эту методику дистрибутивного разграничения лексем можно описать следующим образом: из текста [15] достаточного объема извлекаются все ориентированные фразы, содержащие исследуемое слово в ядерной позиции. Затем все слова фразы за исключением ядра заменяются символами тех функциональных классов, к которым они принадлежат. Таким образом вычленяются дистрибутивные формулы. Полученные формулы сравниваются. Разным формулам приписывается разное значение; идентичные формулы объединяются, и им приписывается одно значение.

---

[14] Эта возможность более или менее подробно была исследована автором в работе «Дистрибутивный анализ значений и структурные семантические поля». «Лексикографический сборник», № 5. (В печати).

[15] В принципе можно исходить не только из текста, но и из словаря. В этом случае меняются лишь технические приемы работы.

Проиллюстрируем сказанное несколькими примерами:

дф[16]		пример
брякать	$N^1_п — N^2_i$	*Она брякает ложками*
	$N^1_п — N^2_a$ *на* V^3_a	*Я брякнул ключ на стол*
	$N^1_п — N^2_d N^3_a$	*Я брякнул ему всю правду*
бряцать	$N_п —$	*Шпоры бряцают*
	$N^1_п —$ *на* N^2_p	*Он бряцает на лире*
	$N^1_п — N^2_i$	*Гусар бряцает шпорами*
бурчать	*У* $N^1_b —$ *в* N^2_p	*У него бурчит в животе*
	$N^1_п — N^2_a$	*Он бурчит что-то*
ввалиться	$N_п —$	*Щеки ввалились*
	$N^1_п —$ *в* N^2_a	*Он ввалился в комнату*

Такие разграничения — не предел возможностей рассматриваемой дистрибутивной методики. Введя новые подразделения в функциональные классы, мы можем получить более точное и глубокое разграничение значений некоторых слов.

Предположим, что нам необходимо определить число лексем для слов *баламутить*, *бегать*, *бренчать* и *будить* в следующих текстах:

Мы баламутим воду	*Чашки бренчат на столе*
Мы баламутим ребят	*Он бренчит на гитаре*
$N^1_п — N^2_a$	$N^1_п —$ *на* N^2_p
Ребята бегают	*Дежурный будит спящих*
Глаза бегают	*Тишина будит мысли*
$N_п —$	$N^1_п — N^2_a$

Если судить по дистрибутивным формулам, в каждой паре текстов представлено по одному значению. Однако интуитивно мы склонны, очевидно, отделять друг от друга тексты во всех четырех примерах как содержащие разные лексемы.

Попытаемся уточнить наше описание. Поскольку в русском языке на структурной основе можно вполне строго выделить категорию одушевленности—неодушевленности, все N могут быть разбиты на два новых класса эквивалентных элементов: N^a (одушевленные) и N^i (неодушевленные). Если подставить в наши формулы новые символы вместо недифференцированного символа N, окажется, что мы можем строго описать интуитивно ощущаемые нами различия, не выходя за рамки чисто дистрибутивной процедуры замещения (субституции).

Мы баламутим воду $N^1_п — N^{2i}_a$	*Чашки бренчат на столе* $N^{1i}_п —$ *на* N^2_p
Мы баламутим ребят $N^1_п — N^{2a}_a$	*Он бренчит на гитаре* $N^{1a}_п —$ *на* N^2_a
Ребята бегают $N^{1a}_п —$	*Дежурный будит спящих* $N^1_п — N^{2a}_a$
Глаза бегают $N^{1i}_п$	*Тишина будит мысли* $N^1_п — N^{2i}_a$

Можно было бы пойти несколько дальше по пути уточнения дистрибутивных формул. Однако есть более эффективное средство усиления

[16] Для краткости в дистрибутивных формулах вместо морфемы ставится черточка. Примеры заимствуются в основном из современных словарей русского языка.

дистрибутивного анализа, связанное с предложенным выше понятием развернутых дистрибутивных формул.

б) Вторая методика.

Если продолжать дистрибутивный анализ значений в том виде, в каком он был изложен в предыдущем параграфе, то рано или поздно мы столкнемся со следующими двумя серьезными трудностями: во-первых, одна и та же дистрибутивная формула скрывает несколько разных значений; во-вторых, разные дистрибутивные формулы реализуют лексемы, которые интуитивно нами отождествляются.

Пусть у нас имеются тексты *группа бежит* и *время бежит* ($N_п$ —). Дистрибутивная формула не вскрывает различия лексем, хотя эти различия интуитивно достаточно ясны. Легко заметить, однако, что во фразе *группа бежит* после слова *бежит* имеется несколько факультативных позиций, ср. *группа бежит обедать*, *группа бежит в столовую* или *группа бежит в столовую обедать*. Следовательно, лежащую в основе этой фразы дистрибутивную формулу можно развернуть и представить так: $N^1_п$ *бежит* (*в* $N^2_а$)(V). Во фразе *время бежит* нет аналогичных позиций. Лежащую в основе этой фразы дистрибутивную формулу нельзя развернуть так, как была развернута предыдущая формула. На основании структурного противопоставления «факультативная позиция» — «минимая позиция» заключаем, что в наших текстах налицо две противопоставленных лексемы (т. е. два разных значения слова *бежать*).

Ср. также фразы *ребята бросались снежками* (*в прохожих*) и *они бросались работниками*, где развертывание фраз выявляет скрытые до того различия и создает структурную основу для разграничения лексем.

Перейдем теперь ко второму из названных выше случаев. Пусть нам даны тексты *он идет* и *он идет в столовую обедать*. Требуется определить, сколько значений слова *идти* представлено в этих текстах. Анализ показывает, что лексема *идет* может быть отождествлена в обеих фразах, поскольку в кратчайшей фразе есть обе факультативные позиции, которые фактически реализуются во второй фразе. Если в любой из данных фраз проведение операций свертывания и развертывания не приводит на каком-либо шаге к нарушению отмеченности фразы, они семантически отождествляются относительно своего ядра. Лексема *идти*, представленная в первых двух фразах, отличается от лексемы *идти*, содержащейся во фразе *дорога идет в Лондон*, поскольку эта фраза не имеет позиции V и не допускает соответствующего развертывания.

Более трудную проблему представляет разграничение лексем во фразах *директор болеет* и *директор болеет за успех дела*. Первая фраза как будто развертывается во вторую, а вторая свертывается к первой. Чтобы определить, законна ли такая операция и не происходит ли в этом случае «переключения» значения, пробуем все возможные развертывания для этих фраз. В первой фразе есть факультативная позиция $N^2_п$, ср. *директор болеет тифом*. В этой новой фразе уже нет пустой позиции *за* $N^3_а$ (ср. *за успех дела*). Во второй фразе тоже имеется факультативная позиция $N^2_і$, которая, однако, совместима с позицией *за* $N^3_а$, ср. *директор болеет душой за успех дела*. В итоге мы получаем для слова *болеет* в наших текстах две дистрибутивных формулы, которые адекватно различают две представленные в них лексемы: $N^1_п$ *болеет* ($N^2_і$) и $N^1_п$ *болеет* ($N^2_і$) *за* $N^3_а$.

Анализ материала показывает следующие интересные зависимости. Свернутые дистрибутивные формулы обладают минимальной диагностирующей силой, т. е. предсказывают максимум значений. Развернутые дистрибутивные формулы обладают максимальной диагностирующей силой, т. е. предсказывают минимум значений. Дистрибутивные формулы, в которых можно свернуть или развернуть какие-нибудь позиции, ди-

агностируют общее (неспециализированное) значение. Дистрибутивные формулы, в которых нельзя ни свернуть, ни развернуть какие-либо позиции, диагностируют специализированное значение.

Поскольку максимальной диагностирующей силой обладают развернутые дистрибутивные формулы, при фиксации той или иной лексемы в словаре целесообразно показывать все факультативные позиции, содержащиеся в типичной для нее дистрибутивной формуле.

Изложенная здесь методика является гораздо более гибким инструментом исследования значений, чем методика, описанная в предыдущем параграфе. Однако и в этом случае мы сталкиваемся на известном этапе анализа с одной из двух серьезных трудностей, о которых мы говорили в начале данного параграфа. Оказывается, что иногда даже развернутая дистрибутивная формула может скрывать две разных лексемы (см. часть I, § 5). Такие случаи могут быть решены только на трансформационном уровне анализа.

2. Трансформационная методика разграничения лексем

Процедуру разграничения лексем на основе трансформационных критериев можно описать следующим образом. Фразы, которым соответствуют разные развернутые дистрибутивные формулы, далее не анализируются (*установленное различие считается релевантным в данном языке*). Остальные фразы подвергаются всем возможным трансформациям (т. е. трансформациям, имеющимся в нашем списке).

Если оказывается, что разные фразы допускают одни и те же трансформации, то ядра, относительно которых они сравниваются, семантически отождествляются. Если оказывается, что разные фразы характеризуются разными трансформациями, ядра, относительно которых они сравниваются, семантически разграничиваются.

Пусть мы отождествили на основе дистрибутивных критериев ядро *пишет* во фразах *он пишет час* и *он пишет письмо* (в данном случае мы берем для удобства частично развернутые фразы, но в принципе в них можно развернуть еще несколько аналогичных позиций, ср. *он пишет другу час* и *он пишет другу письмо* и т. д.). Вторая фраза допускает пассивную трансформацию (*письмо пишется им*), которая невозможна для первой фразы. На основании этого структурного противопоставления устанавливаем два разных значения слова *писать* (транзитивное и нетранзитивное) [17]. Ниже даются иллюстрации.

Швырять в конструкции $N^1_п V N^2_а$ значит «бросать в разных направлениях» [18], если: $N^1_п V N^2_а \rightarrow N^2 V_{от} N^{1\,19}$, ср. *волны швыряли корабль* → *корабль швыряло волнами*. Если такая Т невозможна, *швырять* значит «заставлять лететь», ср. *дети швыряли камни*.

---

[17] Вообще говоря, этот принцип дает возможность определить для любого случая список морфологических элементов (включая сюда и словообразовательные аффиксы) и синтаксических конструкций, которые могут играть существенную роль в разграничении лексем.

[18] Здесь и далее (см. части III и IV) такие содержательные формулировки даются в чисто эвристических целях. Как мы неоднократно подчеркивали в данной работе, предлагаемая нами формальная процедура дает объективный критерий лишь для разграничения значений. Что касается содержательного толкования устанавливаемых категорий, то на данном этапе исследования оно может быть лишь интуитивным. См. также Заключение.

[19] Формула типа: $N^1_п V N^2_а \rightarrow N^2_п V_{ся} N^1_і$ читается следующим образом: конструкция $N^1_п V N^2_а$ под действием некоторой операции преобразуется в конструкцию $N^2_п V_{ся} N^1_і$.

Бросаться в конструкции N^1_nV *в* $N^2_aN^3_i$ значит «устремляться», если $N^1_nV^1$ *в* $N^2_aN^3_i \rightarrow N^1_nV^2$ *в* $N^2_aN^3_i$, ср. *он бросался в драку зверем* → *он бросился в драку зверем*. Если такая Т невозможна, *бросаться* значит «заставлять лететь», ср. *дети бросались в щит снежками*.

Весить в конструкции $N^1_nVN^2_a$ значит «иметь вес», если $N^1_nVN^2_a \rightarrow N^1_nN^3_i$ *в* N^2_a, ср. *рыба весит килограмм* → *рыба весом в килограмм*. Если такая Т невозможна, *весить* значит «взвешивать», ср. *они весили свинью*.

Содержать в конструкции $N^1_nVN^2_a$ значит «иметь в себе», если $N^1_nVN^2_a \rightarrow$ *в* N^1_pV *ся* N^2_n, ср. *письмо содержит намек* → *в письме содержится намек*. Если такая Т невозможна, *содержать* значит «давать средства», ср. *отец содержит семью*.

Предложенная процедура дает возможность определить формальный (структурный) порог для установления релевантных в данном языке различий. Семантические различия, которые могут нами ощущаться, *но не имеют формального (структурного) выражения, являются, очевидно*, случайными. Связанные с узкими, не обобщаемыми контекстами, они практически бесконечны и не должны учитываться в научном лингвистическом описании.

На основании предложенной методики можно, вероятно, преодолеть один из самых существенных недостатков традиционной лексикографии — неумение отличать существенные семантические противопоставления от случайных. Как известно, ни для одного языка не установлена номенклатура лексических значений (как, скажем, установлена номенклатура фонем). Обычно она изменяется от словаря к словарю и в большой мере зависит от объема словаря. Словари большего объема дают как правило более дробную классификацию значений, чем словари меньшего объема. Стремление фиксировать абсолютно все оттенки значений делает задачу лексикографа необозримой и неразрешимой, так как при такой постановке дела описание в принципе стремится к бесконечности. Структурная точка зрения дает возможность сформулировать для лексикографа некоторую конечную задачу и тем самым сделать научным описание лексики (словарь).

III. КЛАССЫ ЭКВИВАЛЕНТНЫХ ЛЕКСЕМ

После того как в рамках установленного словаря (порядка 5—8 тыс. слов) завершено разграничение значений, т. е. получен список лексем, различие между которыми релевантно в данном языке, первый этап работы можно считать законченным. На втором этапе делается попытка объединить полученные лексемы в классы эквивалентных лексем.

Признаком, на основе которого лексемы объединяются в классы эквивалентных лексем, является способность употребляться в одной и той же конструкции или трансформационном комплексе конструкций. Конструкции и трансформационные комплексы конструкций обладают известной диагностической силой, т. е. предсказывают класс эквивалентных лексем, которые характеризуются некоторым семантическим инвариантом. Разные конструкции и трансформационные комплексы конструкций обладают разной диагностической силой, предсказывая от одной до $n-1$ лексемы. Так, конструкция N^1_nV*в*$N^2_a\bar{V}$ обладает меньшей диагностической силой, чем конструкция $N^1_dV_\varnothing$ *от* N^2_g, поскольку позиция V может замещаться бо́льшим числом членов V в первом случае по срав-

нению со вторым; ср. следующие фразы с пропуском глагола: *он — в столовую обедать* (*идет, бежит, едет, мчится, скачет, бредет, торопится, плывет, направляется* и т. д.) и *мне — от отца* (*попало, досталось, влетело, перепало*).

При построении классов эквивалентных лексем можно идти двумя путями. Во-первых, можно выделить некоторое число конструкций и трансформационных комплексов конструкций и приписать всем им некоторое значение, поскольку они являются самостоятельными элементами структуры. Тогда каждая конструкция и каждый трансформационный комплекс конструкций будут символизировать определенный класс эквивалентных лексем. Во-вторых, классы эквивалентных лексем можно строить операционно. Тогда значение будет приписываться не любой конструкции, но двум исходным противопоставленным конструкциям и каждой операции (дистрибутивной или трансформационной), в результате которой исходная конструкция преобразуется в некоторую другую конструкцию. Тогда каждая новая операция будет символизировать два класса эквивалентных лексем.

Рассмотрим подробнее каждую из этих возможностей.

1. Первая методика

Допустим, что в результате дистрибутивного анализа значений мы выяснили важную роль дистрибутивных формул, которым соответствуют следующие конструкции: 1) $N^1_nVN^2_a$, 2) $N^1_nVN^2_d$ *что* S, 3) $N^1_nVN^2_d\bar{V}$, 4) N^1_nV *в* $N^2_a\bar{V}$, 5) *у* $N^1_gVN^2_n$, 6) $N^1_dV_{\varnothing}$ *от* N^2_g. Конструкции расположены в порядке нарастающей диагностической силы. Рассмотрим диагностируемые ими классы эквивалентных лексем.

1. $N^1_nVN^2_a$. В этой конструкции употребляется подавляющее большинство лексем фиксированного словаря, и она почти не обладает диагностической силой, ср. *охотник убивает волка, я люблю отца, он перешел поле, отец думал минуту, гусар скачет вторую милю, туша весила тонну, картофель стоит рубль, письмо содержит намек, лодка вмещает десять человек* и т. д. Мы можем условно приписать этой конструкции какое-то значение самого общего характера, хотя не вполне ясно, каким образом его формулировать.

2. $N^1_nVN^2_d$ *что* S. Данная конструкция диагностирует класс лексем со значением сообщения, ср. *я говорю (сообщаю, передаю, телеграфирую, пишу, телефонирую, радирую) ему, что скоро приеду.*

3. $N^1_nVN^2_d\bar{V}$. Данная конструкция диагностирует класс лексем со значением побуждения, ср. *я велел (предложил, предписал, приказал, посоветовал) ему уйти.*

4. N^1_nV *в* $N^2_a\bar{V}$. Данная конструкция диагностирует класс лексем со значением передвижения, ср. *он бежал (шел, брел, ехал, тащился, плыл, полз) в цирк смотреть представление.*

5. *У* $N^1_gVN^2_n$. Данная конструкция диагностирует класс лексем со значением наличия, ср. *у меня были (имелись, находились, нашлись) часы.*

6. $N^1_dV_{\varnothing}$ *от* N^2_g. Данная конструкция диагностирует весьма узкий класс лексем, ср. *мне попало (влетело, досталось) от отца.* Она обладает почти максимальной диагностической силой, предсказывая лексемы со значением энергичного осуждения.

Разбиение на классы можно было бы уточнить с помощью трансформационных процедур. Так, первый класс эквивалентных лексем допускает разбиение на два новых класса в зависимости от того, как

реагируют фразы с этими лексемами на Т: $N_h^1VN_a^2 \rightarrow N_n^2V$ *ся* N_i^1 или Т: $N_n^1VN_a^2 — N_n^3N_g^2N_i^1$, ср. *ученый открыл радий, радий был открыт ученым* или *открытие радия ученым*. Этот комплекс конструкций отделяет переходные глаголы от непереходных.

Однако изложенное трансформационное уточнение классификации не устраняет ее существенных недостатков, так как они порождаются самой методикой работы.

Эти недостатки суть следующие:

1) Образующиеся классы эквивалентных лексем совершенно неравноправны, поскольку они содержат от одной до n—1 лексем и, следовательно, семантически несопоставимы.

2) Неясно соотношение образующихся классов. В принципе, очевидно, возможны 4 типа отношений: независимость, или несовместимость, пересечение, включение и тождество, или равнозначность, причем нельзя определить, в каких типовых случаях мы имеем дело с тем или иным видом отношений. Иными словами, полученные классы эквивалентных лексем не обнаруживают строгой иерархии.

Эти недостатки можно было бы устранить, введя в модель некоторые статистические критерии [20]. Тогда иерархию классов можно было бы построить на основе критерия частотности той или иной конструкции. Конструкции с одинаковым индексом частотности расположились бы на одном уровне. Каждая конструкция с более высоким индексом частотности подчиняла бы себе несколько конструкций с более низким индексом частотности и т. д.

Таким образом, существенный недостаток изложенной модели устраняется, однако он устраняется ценой уничтожения однородности модели, в которой теперь, наряду со структурными критериями, используются критерии статистические [21].

Очевидно, разумнее попытаться решить возникающие при построении семантической модели трудности, не прибегая к статистике. В связи с этим обратимся ко второй методике.

2. Вторая методика

В данном разделе нам удобно не различать дистрибутивных и трансформационных операций, а говорить об операциях (О) вообще. Под действием некоторых О множество лексем последовательно разбивается на ряд классов.

На каждом этапе анализа применение определенной О к материалу дает нам два класса эквивалентных лексем: класс K_1, члены которого реагируют на нее положительно, и класс K_2, члены которого реагируют на нее отрицательно.

В принципе может быть автоматически определен порядок примене-

---

[20] Ср. упоминавшуюся работу автора «Дистрибутивный анализ значений и структурные семантические поля».

[21] Здесь необходимо сделать одно замечание о соотношении статистических и структурных методов при построении теоретической модели языка. В принципе, очевидно, вполне мыслимо построение теоретической модели без помощи статистики. Однако в большинстве фактически созданных структурной лингвистикой моделей в явной или неявной форме использовалась статистика. Применение статистических методов может быть плодотворным, если строго различать два уровня анализа: уровень языка и уровень речи. На уровне речи необходимо использовать статистические методы. Они могут сыграть решающую роль в выделении некоторого минимума наиболее частых элементов, без знания которых невозможен никакой структурный анализ. Этим статистически выделенным элементам можно придать силу аксиом; затем их можно использовать в качестве аксиом в развиваемой чисто структурными методами теоретической модели. Однако введение статистических критериев на равных основаниях со структурными в самое теоретическую модель не должно допускаться.

ния различных О к материалу: если некоторые О разбивают исходное множество лексем n на два класса, причем один из классов содержит a членов, а другой $n-a$ членов, то номер операции в списке определяется значением p, вычисляемым по формуле

$$p = \frac{a(n-a)}{(n/2)^2}$$

Чем ближе значение p к единице, тем меньше номер операции в списке. Иными словами, список возглавляет О (дизъюнкция конструкций или трансформационных комплексов, если речь идет о первом шаге), разбивающая исходное множество лексем на два класса с равным или близким к равному числом членов в каждом. Список замыкает О, разбивающая данное множество на два таких класса, один из которых содержит n—1 или близкое к этому число лексем, а другой — одну лексему или близкое к единице число лексем. Между этими двумя крайними О располагаются на основе тех же принципов все другие О.

Если же некоторые операции O_1 и O_2 разбивают данное множество лексем на разные, но количественно равные друг другу классы, то иерархия классов может быть построена двумя разными способами, и выбор того или иного способа определяется исключительно соображениями удобства.

В данной работе, построенной на ограниченном, скорее иллюстративном материале, порядок применения О строго не определялся. Мы можем предложить поэтому лишь примерную схему разбиения глагольных лексем на классы эквивалентных лексем [22].

Пусть отправной точкой нашего анализа будет конструкция $N_n^1VN_a^2$, диагностирующая, как было показано выше, весьма широкий К. О: $N_n^1VN_a^2 \rightarrow N_n^2V$ *ся* N_i^1 или О: $N_n^1VN_a^2 \rightarrow N_n^3\ N_g^2\ N_i^1 > $ К 1 со значением транзитивности и К 2 со значением нетранзитивности. К 1: *я читал книгу* → *книга читалась мной, мы абонировали ложу* → *ложа была абонирована нами, войска переходят поле* → *переход поля войсками* и т. д. К 2: *мальчик бежал минуту, гусар скачет вторую милю, туша весит тонну, письмо содержит намек*, и т. д.

1. Анализируется К 1.0: $N_n^1pVN_a^2 \rightarrow N_n^1VN_i^2 >$ К 11 со значением транзитивного движения (субъектного перемещения) и К 12 со значением транзитивных действий и состояний. К 11: *дети перебежали (перешли, переехали, обошли, переползли, проехали) поле* → *дети бежали (шли, ехали, ползли) полем; парусник переплыл океан* → *парусник плыл океаном* и т. д. К 12: *ловить, открывать, резать, вбивать, чувствовать, читать, тянуть, калечить, любить, видеть, держать, обнимать, понимать, слышать* и т. д.

12. Анализируется К 12. 0: $N_n^1VN_a^2 \rightarrow N_n^1VN_d^3N_a^2 >$ К 121 со значением внешне выраженных действий и состояний и К 122 со значением внешне не выраженных действий и состояний [23]. К 121: *я ловил соседу рыбу, он зарядил охотнику ружье, хирург открыл мальчику рану,*

[22] Поскольку в дальнейшем изложении необходимо многократно повторять одни и те же формулировки, принимается следующая достаточно экономная система символической записи операций: фиксируется О, применяемая на данном шаге анализа; затем ставится знак порождения > и фиксируются полученные в результате этой операции К (классы эквивалентных лексем), причем на первом месте стоит К, члены которого допускают положительную реакцию содержащей их фразы на данную О, а на втором месте стоит К, члены которого не допускают положительной реакции содержащей их фразы на данную О. Каждому К приписывается номер (формальное определение), после чего он определяется содержательно. В конце параграфа даются примеры.

[23] Здесь и далее семы предшествующих классов, определяющие и данный класс, для удобства изложения не повторяются.

брат читал мне книгу, мы красили ей платье и т. д. К 122: *ощущать, видеть, обожать, уважать, помнить, знать, понимать, чувствовать, обонять, воспринимать* и т. д.

121. Анализируется К 121.0: $N_п^1VN_а^2 \rightarrow N_п^1VN_а^2$ *на* $N_а^3 >$ К 1211 со значением направленных действий и К 1212 со значением ненаправленных действий и состояний. К 1211: *крестьяне везли овощи на рынок, мы косили траву на сено, портной шил платье на мальчика, они пустили машину на лом, рабочие пилили лес на дрова, рыбаки несли сети на берег, мы учим детей на благо родины* и т. д. К 1212: *он держал кувшин, пожарные гасили огонь, лошади поднимали пыль, они улучшают питание, работница чистит ковер, боец точит штык.*

1211. Анализируется К 1211. 0: $N_п^1VN_а^2 \rightarrow N_п^1VN_а^2$ *на* $N_а^3$ *(для* $N_в^3$*)* $>$ $>$ К 12111 со значением объектно-ориентированных действий и К 12112 со значением объектно-неориентированных действий. К 12111: *шить костюм на/для мальчика, делать работу на/для хозяина, строгать доски на обшивку (для обшивки), убивать птиц на ужин (для ужина), производить машины на продажу (для продажи)* и т. д. К 12112: *везти, ловить, открывать, резать, рубить, убирать, класть, гнать, гнуть, выгибать, тереть, пускать, бить, косить, вести, доставлять, наливать, вбивать, тянуть, рвать* и т. д.

12112. Анализируется К 12112. 0: $N_п^1VN_а^2 \rightarrow N_п^1VN_а^2$ *в/на* $N_а^3 \rightarrow N_п^3V$ *ся* $N_і^1$ *из* $N_в^2 >$ К 121121 со значением превращения и К 121122 со значением перемещения. К 121121: *резать бумагу в куски → куски режутся из бумаги, рубить дерево в щепы → щепы рубятся из дерева, свертывать бумагу в трубочку → трубочка свертывается из бумаги, молоть зерно в муку → мука мелется из зерна, тереть камень в порошок → порошок трется из камня, косить траву на сено → сено косится из травы, тянуть брусок в проволоку → проволока тянется из бруска* и т. д. К 121122: *тащить корзину на базар, тянуть ее на танцплощадку, наливать молоко в стакан, вбивать гвоздь в стену, вести друга в город, везти его в клуб, нести бочку на берег, пускать товар в оборот, складывать яблоки в корзину, слать письмо в город, убирать работу в ящик, гнать стадо в загон, класть яйца в сено, писать письмо на Украину* и т. д.

122. Анализируется К 122. 0: $N_п^1VN_а^2 \rightarrow N_п^1V\bar{V} >$ К 1221 со значением субъективных состояний и действий и К 1222 со значением объективных состояний и действий. К 1221: *он любил выпить, я предпочитаю не видеть этого, сестра обожает обедать на открытом воздухе, он ненавидит ходить туда* и т. д. К 1222: *воспринимать, видеть, слышать, обонять, осязать, слушать, нюхать, ощущать, чувствовать, знать, понимать, помнить* и т. д.

2. Анализируется К 2.0: $N_п^1VN_а^2 \rightarrow N_п^1V$ *в* $N_а^3\bar{V} >$ К 21 со значением движения и К 22 со значением действий и состояний. К 21: *я бежал в столовую обедать, он шел в комнату спать, спортсмен плыл в бассейн смотреть соревнования, ученик входил в библиотеку читать, он бегал в галерею смотреть картины* и т. д. К 22: *стоять, лежать, сидеть, болеть, думать, шутить, мечтать, дрожать, страдать, скучать, жить* и т. д.

21. Анализируется К 21.0: $N_п^1VN_а^2 \rightarrow N_п^1VN_а^2N_а^3 >$ К 211 со значением направленности и К 212 со значением ненаправленности. К 211: *он бежал (шел, плыл, полз) дистанцию час.* К 212 *бегать, ходить, плавать, ползать, бродить, кататься, блуждать* и т. д.

22. Анализируется К 22.0: $N_п^1VN_а^2 \rightarrow N_п^1V >$ К 221 со значением неизмеряемых действий и состояний и К 222 со значением измеряемых действий и состояний. К 221: *я думал (скучал, шутил, мечтал, сидел,*

лежал, стоял, жил, дрожал, болел). К 222: *комната вмещает пять человек, письмо содержит два предложения, телеграмма заключала три слова, стадо насчитывает сто голов, обед стоит рубль, рыба весит килограмм, дом имеет пять комнат* и т. д.

221. Анализируется К 221.0: $N_n^1V \rightarrow N_n^1V\ o\ N_p^2 >$ К 2211 со значением нефизических действий и состояний и К 2212 со значением физических действий и состояний. К 2211: *я думал о тебе, юноша мечтал о славе* и т. д. К 2212: *сидеть, лежать, стоять, дрожать, болеть* и т. д.

222. Анализируется К 222.0: $N_n^1VN_a^2 \rightarrow$ *в* N_p^1V *ся* $N_n^2 >$ К 2221 со значением наличия и К 2222 со значением меры. К 2221: *комната вмещает пять человек → в комнате вмещается пять человек, письмо содержит два предложения → в письме содержится два предложения, телеграмма заключала три слова → в телеграмме заключалось три слова, стадо насчитывает сто голов → в стаде насчитывается сто голов, дом имеет пять комнат → в доме имеется пять комнат* и т. д. К 2222: *стоить, весить* и т. д.

Можно было бы продолжать этот анализ, но цель наша состоит не в том, чтобы построить в данной работе исчерпывающую семантическую классификацию русского глагола, а в том, чтобы показать возможность создания такой классификации на структурной основе и осмыслить принципы ее построения.

3. Иерархия классов

Как мы видели, на основе структурных признаков можно получить иерархическую систему классов эквивалентных лексем, расположенных таким образом, что каждый класс высшего порядка состоит из двух непересекающихся классов низшего порядка. При строгом подходе к делу эту иерархическую систему можно представить в виде дерева с классом эквивалентных лексем в каждом его узле [24].

Оказывается, что если разбиение произведено указанным способом, мы получаем возможность:

1) формально определить каждый класс эквивалентных лексем,

2) формально определить отношения между классами,

3) формально определить расстояние между любыми двумя классами, т. е. точно измерить семантическую близость классов,

4) провести компонентный анализ значений.

Определить каждый класс можно следующим образом. Поскольку узлы дерева символизируют не только класс, но и операцию, в результате которой он получается из предшествующего класса (если данный класс не является исходным множеством лексем), а также свойство класса (сему), — каждый класс может быть определен через набор операций, которые порождают его из предшествующего класса (операционно) или через набор сем, взятых последовательно из всех узлов дерева, к которым этот класс может быть возведен.

Формальной записью определения класса и в том и в другом случае служит запись вида К 1, К 2, К 1211, К 222 и т. п. (букву можно опускать). Запись имеет довольно простой вид. Однако структура такого формального определения является достаточно сложной, так как в записи важны: 1) номера сем или операций (цифры 1 и 2 имеют разное значение); 2) их порядок или место в формальном определении (цифра 2 имеет разные

[24] При наличии более двух независимых, т. е. не сводимых друг к другу исходных конструкций, плоскостное расположение классов эквивалентных лексем первого порядка (получаемых из заданного множества лексем на первом же шаге анализа) может оказаться неудобным, и тогда лучше использовать модель многомерного семантического пространства.

значения в определениях 12 и 112); 3) последовательность цифр, которая им предшествует (выделенная цифра имеет разные значения в определениях 221 и 121). Иными словами, значение цифры меняется, если меняется хотя бы один из этих элементов.

Отношения между классами могут быть формально определены в терминах глоссематической модели Л. Ельмслева [25], т. е. в терминах зависимостей трех видов: взаимозависимости (интерденденции), односторонней зависимости (детерминации) и свободной зависимости (констелляции). Два класса эквивалентных лексем А и В, порождаемых из класса С одной и той же операцией О, находятся в отношении двусторонней зависимости друг к другу (в символической записи: $A \leftrightarrow B$). Каждый из классов А и В, непосредственно порождаемых путем операции О из класса С, находится в отношении односторонней зависимости к классу С (в символической записи: $A \rightarrow C$ и $B \rightarrow C$). Все другие отношения между классами в пределах данного дерева сводятся к отношению свободной зависимости (в символической записи: A I B).

Определение отношений между классами представляет интерес с двух точек зрения. Во-первых, это дает возможность сопоставить классы, казалось бы весьма далекие друг от друга, поскольку некоторые из перечисленных типов отношений имеют важное внутреннее сходство. По Л. Ельмслеву, двусторонняя зависимость сближается с односторонней зависимостью, так как и в том, и в другом случае по крайней мере один из элементов является постоянным. Двусторонняя зависимость сближается со свободной зависимостью, поскольку и в том и в другом случае оба элемента противопоставления или постоянные (двусторонняя зависимость), или переменные (свободная зависимость). Интересно было бы дать содержательную интерпретацию этих закономерностей на материале классов эквивалентных лексем. Во-вторых, определение отношений между классами эквивалентных лексем может служить средством измерения расстояний между классами.

Если считать единицами расстояния шаги (т. е. ветви дерева, отделяющие один класс от другого), то расстояние между классами, которые находятся в отношении односторонней зависимости (детерминации), всегда будет равно одному (максимальная семантическая близость). Расстояние между классами, которые находятся в отношении двусторонней зависимости (интердепенденции), всегда будет равно двум. Расстояние между классами, которые находятся в отношении свободной зависимости (констелляции), будет измеряться числом от двух и больше (минимальная семантическая близость). В каждом конкретном случае его достаточно легко подсчитать.

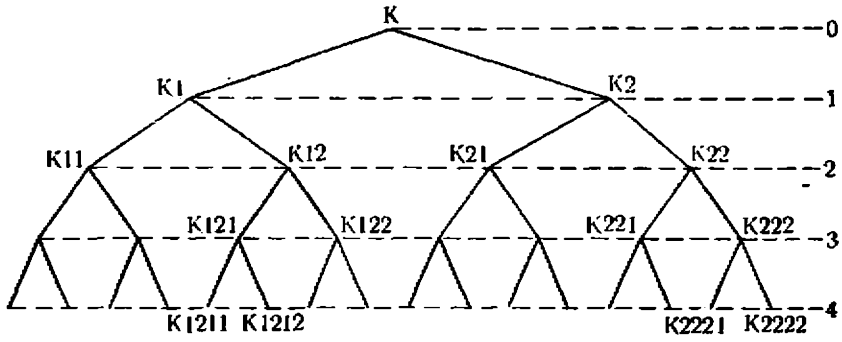
Хотя нужды в какой-либо общей формуле для вычисления расстояний между классами нет, ее в принципе вполне можно сконструировать для всех трех типов отношений. Пусть $S(K_1/K_2)$ будет расстоянием между классами, l_1 — уровень K_1 в дереве, l_2 — уровень K_2 в дереве, а s — уровень общей для них вершины (под уровнем понимается величина, характеризующая степень удаления некоторого узла от вершины дерева). Тогда

$$S(K_1/K_2)=l_1+l_2-2s$$

Ниже дается диаграмма и несколько примеров (см. выше).

---

[25] Ср.: П. А. Соболева. Компонентный анализ значений глагола на основе словообразовательного признака (см. настоящий сборник, стр. 175—189). Однако П. А. Соболева определяет отношения между классами в терминах одновременно двух формальных систем: глоссематики и символической логики.



1) $s(K_1/K2)=1+1-0=2$
2) $s(K121/K122)=3+3-4=2$
3) $s(K121/K1211)=3+4-6=1$ (максимальная семантическая близость)
4) $s(K1211/K2222)=4+4-0=8$

Расстояние между классами можно измерять еще одним способом, исходя не из отношений между классами, а из формальных определений классов (см. стр. 159). Тогда семантическая близость классов вычисляется по формуле

$$S(K_1/K_2)=1-\frac{C}{D}$$

где C — количество сем или операций, общих для K_1 и K_2, а D — количество различных сем или операций. Общая сема или операция — это одна и та же цифра в формальном определении класса, занимающая в этом определении одно и то же место (первое, второе и т. д.) после одной и той же последовательности цифр. Различные семы или операции — это сумма всех различных мест в обоих определениях сравниваемых классов. Эта система измерения семантической близости классов приводит к несколько иным, но вполне сопоставимым с предыдущими результатам.

Сопоставление двух систем измерения семантической близости классов не входит в задачи настоящей работы.

IV. КОМПОНЕНТНЫЙ АНАЛИЗ ЗНАЧЕНИЙ

Вообще говоря, нет особой необходимости подробно излагать методику проведения компонентного анализа значений. Основную работу по компонентному анализу можно считать завершенной после того, как построена иерархия классов эквивалентных лексем и получено формальное определение классов, так как в ходе построения такой иерархии лексемы автоматически разлагаются на составляющие семантические компоненты.

Действительно, все семы, определяющие тот или иной класс, могут быть приписаны любой входящей в него лексеме. Так, лексема *молоть* (ср. *мельник молол зерно*), принадлежащая К 121121, довольно полно определяется набором шести семантических компонентов: транзитивность, действие, выраженность, направленность, объектная неориентированность, превращение (121121). Лексема *гнать* (ср. *пастух гнал скот в деревню*), принадлежащая К 121122, определяется первыми пятью семантическими компонентами из числа только что названных, отличаясь от лексемы *молоть* семантическим компонентом «перемещение» (121122). Таким способом мы можем получить определения всех проанализированных нами лексем.

Однако наши определения не будут исчерпывающими, т. е. не позволят нам отличать друг от друга две лексемы одного и того же конечного класса. Дело в том, что лексемы разлагаются на составляющие их семантические компоненты до известного предела. Крайняя точка анализа задается конечными узлами дерева, из которых не выходят никакие новые ветви. Когда мы достигаем этой крайней точки, компонентный анализ значений в рамках построенной нами иерархии можно считать исчерпанным, хотя до индивидуальной лексемы мы еще не дошли.

Мы получили значение с точностью до класса (может быть, с точностью до синонимически-антонимического ряда), но не получили его с точностью до слова. Существуют, правда, лексемы типа *делать*, *иметь*, *брать*, *давать*, *двигаться*, *действовать* и т. п., содержание которых будет исчерпано набором семантических компонентов, характеризующих определенный класс иерархии. Однако содержание многих лексем, таких как *автоматизировать*, *акклиматизировать*, *амнистировать* и т. п., окажется богаче. Вопрос о том, как нетривиальным образом решить задачу исчерпывающего описания таких лексем (если она вообще может быть решена методами структурной лингвистики, т. е. если различие между такими лексемами релевантно в данном языке), должен стать предметом самостоятельного исследования.

V. ЗАКЛЮЧЕНИЕ

Изложенная формальная модель была применена пока к ограниченному материалу — 400 глаголам русского языка. В принципе она, видимо, может быть использована для анализа слов любого грамматического класса, в том числе и существительных. Правда, существительные нельзя будет разбить на лексемы таким образом, чтобы каждой лексеме (т. е. лексическому инварианту) соответствовало одно «словарное» значение, в то время как между глагольными лексемами и «словарными» значениями глаголов можно в большинстве случаев установить взаимно-однозначное соответствие. Это, может быть, свидетельствует о том, что существительные и глаголы по-разному строят свое содержание. В частности, для существительных принципиальной оказывается функция денотации.

В изложенном понимании структурная лексикология примыкает к структурной грамматике, вместе с которой она образует формальную теорию структуры плана содержания.

Формальные грамматические и лексикологические модели должны быть интерпретированы на семантическом уровне. Иными словами, единицы структуры содержания должны быть соотнесены с единицами семантической субстанции, подобно тому, как фонологические единицы связываются с фонетическими. Для выполнения этой задачи необходимо создать объективную эмпирическую теорию смысла.

ON THE NOTIONS AND METHODS OF STRUCTURAL LEXICOLOGY

Summary

This paper furnishes an apparatus for a formal description of the lexical morphemes of a language, by means of distributional and transformational procedures. All the notions essential to the study are deduced from a set of assumptions. The resultant formal model provides a basis for (1) setting up semantic distinctions relevant in the given language, (2) establishing classes of equivalent lexemes and measuring distances among them, (3) analysing lexemes into elementary semantic components.

В. И. Перебейнос

К ВОПРОСУ ОБ ИСПОЛЬЗОВАНИИ СТРУКТУРНЫХ МЕТОДОВ В ЛЕКСИКОЛОГИИ

Прогресс всякой науки в значительной мере зависит от строгости и точности определения ее основных понятий, а также от степени объективности применяющихся в ней методов исследования. Положение языкознания с этой точки зрения весьма неприглядно, так как, по высказыванию Ф. де Соссюра, «до сих пор в области языка всегда довольствовались операциями над единицами, как следует не определенными» [1]. Хуже всего обстоит дело в семасиологии, которая, по всеобщему признанию, значительно отстает от других областей науки о языке.

Определение значения при помощи используемых в семасиологии традиционных логико-семантических методов не удовлетворяет основным требованиям, предъявляемым к определению как логической операции, например, требованиям о недопустимости круга, об однозначности и ясности словесного выражения определения, о недопустимости определять неизвестное через неизвестное.

Определение и анализ значений в семасиологии до сих пор базируются на категориях логики — понятиях, а не на тщательном исследовании языковых форм, использующихся для выражения каждого отдельного значения лексической единицы. Это ведет к крайней субъективности в исследованиях, вследствие чего результаты, получаемые разными исследователями на одном и том же материале, во многом не совпадают.

Давно назрела необходимость найти максимально объективные формализованные методы изучения лексического значения, которые определяли бы значение через языковые элементы и отношения между ними. В работах по структурной лингвистике как в СССР, так и за рубежом неоднократно высказывалась мысль о возможности использования в семасиологических исследованиях структурных методов (в СССР — С. К. Шаумян, В.В. Иванов, Ю. Д. Апресян, И. А. Мельчук и др.; за рубежом Ч. К. Фриз, З. С. Харрис, Н. Хомский и др.).

В настоящей статье излагается попытка применить методы дистрибутивного и трансформационного анализа и метод непосредственно составляющих для разграничения значений многозначного глагола, проследить, до каких пределов можно с их помощью дифференцировать различные значения слова, и установить иерархию этих значений.

Исследование основывается на следующих теоретических предпосылках:

1. В каждой языковой единице диалектически взаимодействуют ее форма и значение: не может быть языковой формы без значения, всякое значение о б я з а т е л ь н о облекается в языке в определенную форму.

[1] Ф. де С о с с ю р. Курс общей лингвистики. М., 1933, стр. 111.

2. Всякое отношение языковых единиц является формой, используемой для выражения определенного значения.

Определенным образом относящиеся языковые единицы могут образовать языковую единицу высшего порядка, т. е. новую форму, выражающую новое значение.

3. В языке существует иерархия языковых форм и соответствующая иерархия выражаемых ими значений.

4. Всякое высказывание (акт речи) является совокупностью языковых форм, служащих для выражения нужного значения.

5. Система формальных средств языка и система выражаемых ими значений находятся в диалектическом единстве.

Исследование проводилось в синхронном плане на материале современного английского языка в следующем объеме [2]:

	Количество источников	Общий объем (в тыс. слов)
драматургия	— 12	261
проза	— 5	244
поэзия	— 20	220
газеты	— 13	270
публицистика	— 11	230
лингвистические тексты	— 9	250
математические тексты	— 7	209
физико-технические тексты	— 12	275
Всего	89	1959

Основываясь на опыте, имеющемся в области машинного перевода, и на высказываниях ряда советских и зарубежных ученых, было принято априори, что текст объемом в одно предложение (от точки до точки) дает достаточную информацию для выявления каждого отдельного значения исследуемого глагола [3]. Каждое такое предложение было закодировано [4], т. е. представлено в виде цепочки членов функциональных классов и групп слов, каждый из которых имеет определенную характеристику его формы, что выражается соответствующей формулой.

С целью придания формулам более компактного и обобщенного вида производится их свертывание по правилам метода непосредственно составляющих: каждая эндоцентрическая конструкция обозначается через ее основной элемент. Например, *a very old man* (формула AD 31) — *man* (формула 1) *highly interesting* (D3) — *interesting* (3), *shall make* (B 2) — *make* (2), *has made* (B 2d) — *make* (2), *was made* (Bbd 2d) — *made* (2p, т. е. глагол в пассиве) и т. д.

Полученные таким образом формулы текстов подвергаются анализу с целью установления характерных для глагола *make* дистрибуций, оформляющих разные его значения.

Чтобы определить минимальный по объему контекст, необходимый для выявления значения глагола *make*, располагаем все тексты в порядке

[2] При подсчете количества слов в источнике за слово принималась любая буква или последовательность букв между двумя пропусками (пробелами). Такое определение слова принято в машинном переводе. Оно удобно для работы, хоть и не учитывает всех проблем, связанных с определением слова в лексикологии.

[3] Количество предложений, недостаточных для выявления значения *make*, составило 1,4%.

[4] Используется в несколько измененном виде код, предложенный Ч. Фризом (The Structure of English). Члены функциональных классов слов: 1 — имя существительное и другие части речи в функции существительного, 2 — полнозначные глаголы, 3 — прилагательные и другие части речи в функции прилагательного, 4 — наречия; функциональные группы слов: A — артикль и слова, которые могут стоять на месте артикля, B — вспомогательные и модальные глаголы, F — предлоги, D — определители к третьему функциональному классу слов и т. д.

нарастающей длины формул. Самыми короткими оказываются тексты, в которых значение *make* детерминируются одним словом, принадлежащим к классу 1 или 3 или к группе F. Кроме этого, как показывает материал, детерминировать значения *make* могут также сочетания 1F, F1, 13, 31, 12, 1F1, FF 1, стоящие справа от него.

Подвергнем анализу те тексты, в которых после *make* следует представитель класса 1.

Глагол *make*, как и всякий другой глагол в английском языке, должен входить по крайней мере в одну из характерных для глагола типов конструкций, а именно: конструкция «деятель—действие» (кодируется нами 12) или конструкция «действие—объект действия» (кодируется 21). Как видно из формул самых кратких текстов, характерных для *make*, этот глагол может входить либо в обе эти конструкции сразу, образуя трехчленную формулу 121, либо только во вторую конструкцию — тогда получаем двухчленную формулу 21. Но он не может входить только в первую конструкцию 12. Очевидно, элемент, обозначающий субъект действия, в ряде случаев не играет роли для выявления значения *make* и может быть опущен (иметь нулевую форму, так как его можно восстановить по окружению данного текста). Это дает нам право сказать, что 21 является вариантом 121.

Таким образом, мы имеем возможность выделить конструктивную модель № 1, описываемую формулой 121 и ее вариантом 21 и детерминирующую значение *make* № 1. Анализ материала показывает, что эта модель в текстах может иметь разные варианты структурного оформления и что в ряде случаев для ее выделения необходимо провести в текстах ряд трансформаций, которые задаем тремя списками [5]:

1) трансформации сложного предложения — 12 трансформаций,
2) трансформации простого предложения — 27 »
3) трансформации словосочетаний — 20 »

После проведения соответствующих трансформаций относим к модели № 1 все тексты, в которых после исследуемого глагола стоит слово класса 1, являющееся последним в трансформированной части текста.

Все оставшиеся тексты можно разделить на следующие группы, принимая во внимание элементы после *make*:

1) 211, 21F, 21F1F1 и т. д., т. е. с нарастанием длины формулы;
2) 213, 213F1, 213 *that* S (предложение после *that*), 213 21F1 и т. д.;
3) 212, 2121, 2121 F1, 212 F 1, 2121 F1F1 и т. д.;
4) 23, 231, 23 F1, 23 *that* S и т. д.;
5) 2 F, 2 F1, 2F1F1 и т. д.; 2FF1, 2F *that* S, 2F1 S и т. д.

Как видим, в большей части текстов к сочетанию 21 прибавляется еще один или несколько элементов. Дальнейший анализ каждой из указанных последовательностей должен показать, существенны эти элементы для оформления лексического значения *make* или нет.

Анализ последовательности 1211 показывает, что она может быть двух видов: $1^{a}\ 2^{a}\ 1^{в}\ 1^{в6}$ - и $1^{a}\ 2^{a}\ 1^{в}\ 1^{c}$. В первой формуле оба слова, следующие за глаголом, имеют один и тот же индекс, что свидетельствует о их тесной связи друг с другом, так как они соотносятся с одним и тем же объектом действительности, а значит, и о связи каждого из них с глаголом.

---

[5] Большинство применяющихся нами трансформаций описывается в работах: Z. S. Harris. Discourse Analysis. «Language», vol. XXVIII, № 1, 1952; его же: Co-occurrence and Transformation in Linguistic Structure. «Language», vol. XXXIII, № 3, 1957; N. Chomsky. Syntactic Structures. 's-Gravenhage, 1957.

[6] Одинаковые буквенные индексы при знаке 1 показывают, что обозначаемые этим знаком слова могут быть поставлены в формулу 1 *is* 1, т. е. относятся к одному и тому же объекту действительности. Буквенные индексы при знаке 2 указывают на один и тот же или разные глаголы. Для удобства анализа глаголу *make* всегда присваиваем индекс *a*.

Во второй формуле индексы у элементов, следующих за глаголом, разные. Для этой последовательности можно применить трансформацию $1^a2^a1^B1^c \rightarrow 1^a2^a1^c$ *for(to)* 1^B, что позволяет свести данную последовательность к формуле $1^a2^a1^c$ F 1^B. Правда, первую последовательность тоже можно свести к формуле $1^a2^a1^B$ F 1^B (*This practice makes the boy a swimmer* → *This practice makes a swimmer of the boy*). Но существенной разницей между этими двумя последовательностями является то, что 1^B1^B первой последовательности можно подставить в формулу 1 *is* 1 (*the boy is a swimmer*), тогда как сделать это с 1^B1^c нельзя (*He made the boy a pipe* → *He made a pipe for the boy.* Но нельзя сказать: *The boy is a pipe*). Разная сумма возможных трансформаций свидетельствует о разных отношениях между элементами данных последовательностей, а значит, и о разных значениях *make.*

Кроме того, во второй последовательности можно произвести редукцию, опустив 1^B, что не влияет существенно на смысл высказывания. Опущение же одного из 1^B в первой последовательности либо вообще невозможно, либо существенно меняет смысл высказывания. Сравним: *He made the boy a pipe* → *He made a pipe for the boy* → *He made a pipe. He made her a lady* → *He made a lady of her.* Но нельзя сказать: *He made a lady.* Или: *He made him a bishop* → *He made a bishop of him.* Но в *He made a bishop* смысл меняется.

Исходя из всего вышеизложенного, принимаем последовательность $1^a2^a1^B1^B$ за модель № 2, оформляющую значение *make* № 2.

Представляет интерес последовательность $1^a2^a1^B$ F с вариантом 2^a1^B F, где 1^a, обозначающий субъект действия, имеет нулевую форму (*I can't make him out. You must make yourself over. Try to make matters up. He sang his verses as he made them up*).

Ясно, что F, стоящее после 1, не может быть с ним связано. Известно, что F принадлежит к одной из групп функциональных слов, являющихся служебными элементами в языке, и самостоятельно не употребляющихся. Остается предположить, что F связано с глаголом и что конструкция 1^a 2^B 1^a F представляет собою отдельную модель № 3, оформляющую значение *make* № 3. Интересно, что в этой модели могут употребляться немногие F (*out, over, up*). Наличие слова класса 1 при этом обязательно (сравним *They can make a man over*, но не *They can make over*).

Последовательность 121F1 может иметь две разновидности: $1^a2^a1^BF1^c$, и $1^a2^a1^BF1^B$, которая трансформируется в $1^a2^a1^B1^B$ (с переменой местами двух последних элементов в некоторых случаях, что зависит от употребленного предлога): *I made a lady of you* → *I made you a lady. He made water into wine* → *He made water wine.* Данная трансформация сводит $1^B2^a1^BF1^B$ к модели № 2. Другая разновидность данной последовательности $1^a2^a1^BF1^a$ не может подвергаться указанной трансформации. Для этого случая выделяем и отдельно анализируем 1^BF1^c.

Исследуемый материал дал возможность выделить 15 групп 1^BF1^c в зависимости от способа анализа этой последовательности. Наибольший интерес представляют следующие случаи:

1) $1^BF1^c \rightarrow 31^B$: *a hat of wool* → *a woolen hat*;

2) $F1^c$ может заменяться словами *there, then, thus,* а затем редуцироваться без существенного изменения смысла высказывания: *We made an attack beyond V.* → *We made an attack there* → *We made an attack.*

3) Если 1^B имеет глагольную основу, то после него сохраняется то F, которое характерно и для соответствующего непереходного глагола: *We made a run along the shore* → *We ran along the shore.*

Если же 1^B образовано от основы переходного глагола, то после него появляется F (большей частью *of*): *She has never made use of her power* → *She has never used her power.*

4) $F1^c$ представляет собой устойчивое сочетание и не относится к предыдущему $1^в$: *in the name of, in addition to, on the basis of.*

5) $21^вF1^c \rightarrow 1^c$ *is used* $2 = 1^в$: *They made gowns of these trees* → *The trees were used to make gowns.*

6) В последовательности $1^вF1^c$ элемент, для которого характерно сочетание с *make*, — не $1^в$, а 1^c. Последовательностей такого рода не очень много, в качестве $1^в$ выступает ограниченное число представителей класса 1. В соответствии с правилом: если элемент А встречается в окружении X, а для элемента В это окружение не характерно, то в сочетании ВА основным элементом является A^7, — имеем право обозначить последовательность $1^вF1^c$ через 1^c (*a kind of statement* → *statement*).

Во всех этих случаях последовательность $1^вF1^c$ сводится либо к $1^в$, либо к 1^c, а вся конструкция $1^в2^м1^вF1^c$ к $1^в2^м1^в$, т. е. к модели № 1.

Таким образом, оказывается, что последовательность 121F1 в большинстве случаев сводится к 121, т. е. что группа 1F1 бывает существенной для выявления значения глагола *make* только в том случае, когда буквенные индексы знака 1 и справа и слева от F одинаковы.

Естественно предположить, что если в 121F1 группа F1 в большинстве случаев несущественна для оформления значения *make*, то тем более это можно сказать о последующих F1, если они имеются в тексте. Действительно, часть из них можно заменить на 3 по указанной выше трансформации и обозначить затем соответствующую последовательность 31 через 1. Другая часть F1 служит эквивалентом слов *there*, *then*, *thus* и может быть опущена без всякого ущерба для выявления значения *make*. Проведенный эксперимент с определением значения *make* в усеченных таким образом текстах показал, что эти группы несущественны для детерминации значения *make*.

Специфическое построение последовательности 1213 (класс 3 обычно стоит не после слова класса 1, а перед ним или же после глагола 2) заставляет предположить, что это самостоятельная модель, не сводимая к 121. Это подтверждается также и тем фактом, что в текстах с данной последовательностью весьма часто встречаются такие слова класса 1, которые не встречались в 121 (*My presence made Earnest happy*, но не *My presence made Earnest*). Принимаем эту последовательность за модель № 4, которая оформляет значение *make* № 4.

В последовательности 1213 (*that*) S перед 3 обычно стоит проморфема *it*, которая, как известно, употребляется только для замены какого-нибудь другого элемента последовательности. В данном случае *it* употребляется вместо последующего *that* S, так что 1213 *that* S → 12 *that* S 3, т. е. это разновидность предыдущей модели с той разницей, что вместо элемента 1 употреблена целая последовательность, которую З. Харрис называет номинализированным предложением[8] и которая может быть заменена проморфемой класса 1 (*Let us make it plain* (*beyond doubt*) *that we will tolerate no more wars*).

В последовательности 1213F1 группа F1 либо заменяется словами *thus*, *there*, *then*, либо относится к 3, что можно подтвердить тем фактом, что она всегда характерна для данного 3 в любом контексте (*aware of*, *sure of*, *equal to*). В обоих случаях F1 не играет роли в оформлении значения *make*, так что последовательность 1213F1 также сводится к модели № 4. То же можно сказать о других последовательностях, где есть слова, следующие за 3 и либо являющиеся характерным для 3 окружением,

---

[7] Это правило сформулировано З. С. Харрисом в работе: Distributiona Structure, «Word», 1954, vol. X, №. 2.

[8] Z. S. Harris. Co-occurrence and Transformation in Linguistic Structure. «Language», vol. XXXIII, № 3, 1957.

либо подставляемые на место 1 перед 3, которое является проморфемой как в *The tabular arrangement makes it possible to inspect this pattern.*

Последовательность 1212 несомненно представляет собою самостоятельную модель, о чем свидетельствует несколько факторов:

а) 1, стоящее после *make* в этой последовательности, часто не может употребляться в модели № 1 (*Try to make Celia understand. You make my blood run cold*);

б) последовательность 1212 не может трансформироваться по правилу $1^{a}2^{a}1^{в}2^{в} \rightarrow 1^{a}2^{a}1^{в} + 1^{в}2^{в}$, как в *I saw the car stall on the tracks* → *I saw the car + The car stalled on the tracks*, что свидетельствует о тесной связи между $1^{в}$ и $2^{в}$;

в) данная последовательность трансформируется как $1^{a}2^{a}1^{в}2^{в} \rightarrow 1^{a}2^{a}(1^{в}2^{в}) + 1^{в}2^{в}$: *The fear made the girl scream* → *The fear made the girl scream + The girl screamed*[9]; и в этом случае $1^{в}2^{в}$ рассматривается как одно целое.

Данную последовательность считаем моделью № 5, которая оформляет значение *make* № 5.

Все элементы, которые встречаются после $2^{в}$ (1, F1, 1F1, 1F1F1 и пр.), как правило, относятся к этому 2 как характерное для него окружение и не влияют на оформление значения *make* (*Make her give you the change*).

Необходимо выделить в отдельную модель также последовательность 123 и ее вариант 23, так как не вызывает сомнения тот факт, что именно 3 оформляет лексическое значение глагола *make* (т. е. значение № 6, детерминируемое моделью № 6); *I shall make good somewhere. Make ready. Make certain.*

Последовательность 1231 на первый взгляд кажется разновидностью 1213 (модель № 4), однако более детальный анализ ее показывает, что это не так. Лишь немногие члены класса 3 могут употребляться в 1231, так что трансформация 1213 → 1231 редко возможна. Некоторые 3, входящие в последовательность 1231, могут употребляться также и в модели № 6 (123), а это свидетельствует, что 3 в конструкции 1231 важнее для оформления значения *make*, чем в модели 1213. Поэтому принимаем последовательность 1231 за модель № 7, которая оформляет значение *make* № 7.

В последовательности 123F1 группа F1 относится к 3 как характерное для него окружение так же, как это было в последовательности 1213F1. Поэтому относим 123F1 к модели № 6 (*She responded and I made sure of her*).

Последовательность 123 (*that*) S является разновидностью модели № 7, в которой вместо элемента 1 употреблено номинализированное предложение (*Mr. L. Henderson made clear that the U. S. is planning interference in the Arab countries* и *The trial will make clear the importance of closure*).

Последовательность 12F может иметь вариант 2F, например: *Make off. Gold seekers made in before night came.* Не может быть сомнения в том, что именно F оформляет лексическое значение *make* в этих случаях и что мы имеем право выделить эти последовательности в модель № 8, детерминирующую значение *make* № 8. Характерно, что в этой модели употребляются не те F, которые встречались в модели № 3, а другие, такие как *off*, *in*, *away*. Другая особенность модели № 8 по сравнению с моделью № 3 заключается в том, что для детерминирования значения глагола тут не требуется наличие слова класса 1, тогда как в модели № 3 оно обязательно.

Часть оставшихся употреблений *make* содержит такие случаи, когда непосредственно после него следует F1. Последовательность 12F1 имеет две разновидности:

а) в одних случаях она подлежит трансформации 12F1 → 121F [10], т. е. сводится к модели № 3. В этих случаях F после *make* связано не с по-

---

[9] Пример взят из упоминавшейся выше работы З. Харриса «Co-occurrence and Transformation in Linguistic Structure».

[10] Если слово класса 1 заменено проморфемой, то эта трансформация обязательна.

следующим существительным, а с самим глаголом, и при делении на непосредственно составляющие 2F1 членим не как 2+F1, а как 2F+1.

б) Другая часть последовательностей 12F1 не может быть трансформирована подобным образом, и 2F1 делится на непосредственно составляющие как 2+F1, поэтому значение *make* в таких случаях определяется группой F1. Эту часть последовательностей 12F1 выделяем в самостоятельную модель № 9 со значением *make* № 9: *She made for the door*. Как видим, F в этой модели не те, которые могут быть в моделях № 3 и 8 (*for*, *towards*).

Последовательность 12F (*that*) S является вариантом модели № 3. Ее можно трансформировать следующим образом: 12F (*that*) S → 12F1 → 121F: *I can't make out what he writes* → *I can't make out his writing* → *I can't make it out*.

Случаи, включающие последовательность 12FF1, можно разделить на две группы:

а) при делении последовательности на 12F+F1 в первой части оказываются 12F, входящие в модель № 8. Группа F1 может быть заменена словами *then*, *there*, *thus* и не влияет на значение 2F: *Everybody made in before 8 o'clock*.

б) При таком же членении данной последовательности в других случаях оказывается, что в первой части содержатся 2F, входящие в модель № 3. Конечно, сочетание 2F с F1 вместо сочетания со словом класса 1 не может не влиять на значение глагола. Поэтому выделяем эти случаи в самостоятельную модель № 10: *We must make up for lost time. You turn round and make up to me now*.

Установленные таким образом характерные для *make* модели охватывают до 97% всех примеров с *make* и имеют следующую частотность:

модель № 1 — 121 и вариант 21	— 53, 6%
модель № 2 — 1211	— 7, 8%
модель № 3 — 121F	— 3, 8%
модель № 4 — 1213	— 16, 2%
модель № 5 — 1212	— 11, 4%
модель № 6 — 123 и вариант 23	— 1, 6%
модель № 7 — 1231	— 1, 6%
модель № 8 — 12F и вариант 2F	— 0, 1%
модель № 9 — 12F1	0, 5%
модель № 10 — 12FF1	— 0,26%

Каждая из выделенных нами конструктивных моделей — это не отдельная дистрибуция глагола *make*, а ряд дистрибуций, объединяемых внутри одной синтаксической конструкции (тип дистрибуции), и детерминируемое каждой моделью значение носит довольно обобщенный, неспециализированный характер. Степень обобщенности значения зависит от широты сферы употребления данной модели в соответствии с семиологическим законом употребления и содержания знаков: чем шире сфера употребления знака, тем менее специально (тем беднее) его содержание [11].

Этот закон является базой для определения иерархии языковых единиц и соответствующих значений, что необходимо для систематизированного описания значений слова. Структурная лингвистика считает вторичным (производным) такое слово, которое имеет более узкую сферу употребления. Это же положение действительно и для единиц синтагматической оси: конструкция, имеющая более узкую сферу употребления, считается вторичной (производной) при сопоставлении с конструкцией, имеющей более широкую сферу употребления. При этом чем шире сфера

[11] См.: С. К. Шаумян. Структурная лингвистика как имманентная теория языка. М., 1958.

употребления конструкции, тем проще ее строение, ее внутренняя структура; такая конструкция, как правило, состоит из первичных, т. е. непроизводных элементов.

Таким образом, можно наметить следующие критерии для определения иерархии выделенных нами конструктивных моделей и детерминируемых ими значений:

1) широта сферы употребления модели.
2) степень сложности ее структуры,
3) способность данной модели к словообразованию.

Исходя из этих критериев, легко установить, что модель № 1 первична по отношению к другим моделям, которые характерны для *make*, так как, во-первых, сфера ее употребления отличается наибольшей широтой, во-вторых, она проста по структуре, в-третьих, именно эта модель дает почти все производные от *make* (например, *make* (n), *maker*, *making*, *makeless*, и только немногие, как *makebelieve*, *make-ready*, *merry-making* — не от этой модели).

Это значит, что и оформляемое моделью № 1 значение первично. Но в силу большой широты употребления модели № 1 детерминируемое ею значение носит наиболее обобщенный и неспециализированный характер, оно близко к грамматическому значению переходности. Поэтому для дальнейшей дифференциации значений *make*, детерминируемых моделью № 1, необходимо подвергнуть эту модель дальнейшему анализу. Для этого надо снова развернуть формулу 121, обращая внимание на определители перед 1 и на его индексы (+, —, *he*, *it*, *he/it* и т. д.) и группируя тексты, в которых 121 имеет одинаковую дистрибуцию. Всего путем чистой комбинаторики можно выделить более 30 дистрибуций глагола *make*, например:

1) $I^{a}_{\overline{he}}\ 2\ I^{a}_{\overline{he}}$ — *He will make a good doctor.*
2) $I^{1}_{+\,he}\ 2\ I_{+\,he}$ — *Boys ought to make useful farmers.*
3) $I^{a}_{\overline{it}}\ 2\ I^{a}_{\overline{it}}$ — *This line makes the third side of our triangle.*
4) $I^{a}_{+\,he}\ 2\ I^{a}_{\overline{it}}$ — *We'll make a good double.*
5) $I^{a}_{+\,it}\ 2\ I^{a}_{+\,it}$ 5 *thousand meters make 5 kilometers.*
6) $I^{a}_{+\,it}\ 2\ I^{a}_{\overline{it}}$ — *Twelve inches make one foot.*
7) $I^{a}_{\overline{he}}\ 2\ I^{a}_{\overline{he}}$ — *Every convert you make is a vote against war.*
8) $I^{a}_{+\,he}\ 2\ I^{a}_{\overline{he}}$ — *They have broken the mold when they made her.*
9) $I^{a}_{+\,he}\ 2\ I^{a}_{+\,he}$ — *We have made six readers in this factory.*
10) $I^{a}_{\overline{he}}\ 2\ I^{a}_{\overline{it}}$ — *He made a loud remark.*
11) $I^{a}_{\overline{it}}\ 2\ I^{a}_{\overline{it}}$ — *The metal makes contact with the form.*
12) $I^{a}_{\overline{it\ it/the\ y}}\ 2\ I^{a}$ — *We can't make a new Cabinet before Magnus comes in.*

13) $1^{a}\ 2\ 1^{a}$ — *Good fences make good neighbours.*
+ +
it he

14) 2 1 — *An attack was made last Friday on an Asian family.*
−
it

15) 2 1 — *Attempts were made to trace the origin of particular words.*
+
it

Релевантно ли различение всех этих дистрибуций для дифференциации значений *make* в этой модели? Анализ материала показывает, что это не так.

Рассматривая 1, стоящее после *make* в 121, как графему, группируем все тексты по признаку наличия этой графемы (по алфавиту графем). Видим, что некоторые 121 наблюдаются не в одной, а в нескольких (иногда в большинстве) отмеченных дистрибуциях или же могут трансформироваться в несколько дистрибуций. Другие же 121, наоборот, могут иметь только одну дистрибуцию.

В некоторых 121 буквенные знаки индекса 1 и справа, и слева от *make* одинаковы, в других — разные ($1^{a}2\ 1^{a}$ или $1^{a}\ 2\ 1^{b}$).

Некоторые элементы 1 справа от *make* могут употребляться с любым определителем (*a*, *the*, нулевой артикль, местоимение), другие же употребляются только с некоторыми или с одним из них.

Знак 1 может обозначать слова, принадлежащие к разным группам, в зависимости от индекса, стоящего под ним: а) *he*, б) *he/it*, в) *it*, г) *it/they*. Одни из них могут иметь форму и единственного и множественного числа, другие — только одну из этих форм.

Наконец в модели 121 можно проводить различные трансформации:

1. актив ↔ пассив,
2. $21 \rightarrow 2_{(1)}$ — *to make an analysis* → *to analyse*.
3. 121 → 1 *is* 1 — *She makes a good wife* → *She is a good wife*.

Анализ сочетаний *make* с 850 существительными показывает, что разграничивать значения *make* в данной модели следует не по одному какому-нибудь признаку, т. е. не по одной дистрибуции или трансформации, а по совокупности формальных признаков, характерных для каждого отдельного сочетания, а именно, по совокупности возможных для данного сочетания дистрибуций, подвергающихся совокупности возможных для него трансформаций.

Если при разграничении значений, детерминируемых разными моделями, главную роль играли дистрибуции, то при разграничении значений внутри модели главные разбиения осуществляются по линии трансформаций, а потом внутри групп, подлежащих данной трансформации (или совокупности трансформаций), выделяются подгруппы с одинаковыми совокупностями дистрибуций.

Путем последовательных разбиений выделяем следующие группы сочетаний по признаку присущих им трансформаций:

Группа	Трансформации		
	Актив↔пассив	21→2 (1)	121→1 *is* 1
I	−	−	+
II	+	+	−
III	−	+	−
IV	+	−	−
V	−	−	−

В каждой из этих групп можно выделить по нескольку подгрупп с одинаковыми совокупностями дистрибуций.

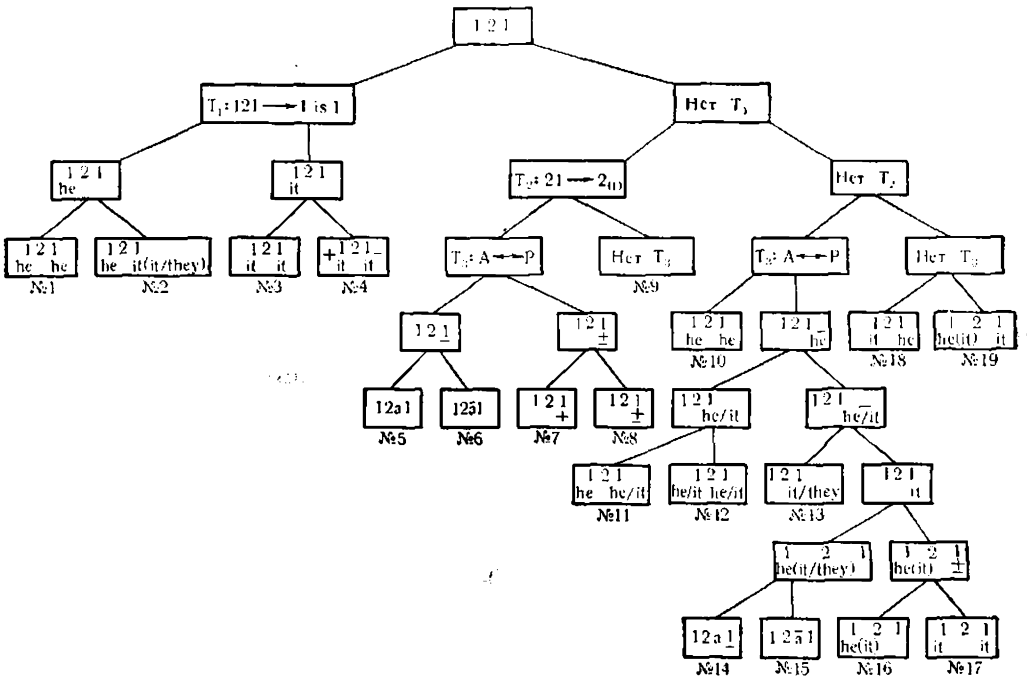
Внутри группы I, например, можно выделить следующие подгруппы:

а) I^a 2 I^a — *He'll make an announcer. Boys ought to make useful farmers.*
∓ ∓
he he

б) I^a 2 I^a — *What a charming little group you've made here.*
∓ ∓
he it/they

в) I^a 2 I^a — *40 rods make a furlong.*
\+ −
it it

г) I^a 2 I^a — *The bush made a good shield. 200 more Hurri-canes made 440 fighters in all.*
\+ ∓
it it



Внутри группы II различаем пять подгрупп. Наибольшее количество подгрупп (более десяти) наблюдается в IV группе. Всего же на основе формальных критериев, упоминавшихся выше, в модели № 1 разграничивается около двадцати значений глагола *make*, как это показано на схеме:

Нет сомнения, что более тщательный анализ может дать возможность выделить более мелкие подгруппы, детерминирующие более специализированные значения глагола. Но уже сейчас ясно, что значение *make* в модели № 1 определяется не семантикой сочетающегося с ним слова, а совокупностью дистрибуций и трансформаций, характерных для каждого отдельного сочетания. Это подтверждается тем фактом, что глагол в сочетании с одним и тем же существительным имеет разные значения в зависимости от указанных выше критериев разграничения значений внутри модели № 1. Например: *He made 60 dollars* (трансформация актив ⟷ ⟷ пассив) и *Your commission makes 60 dollars* (трансформация 121 ↔ 1 *is* 1). Отсюда следует вывод, что значение многозначного слова является функцией характерных для него дистрибуций и трансформаций.

Значения, выделенные таким образом в модели № 1, охватывают до 90% всех случаев употребления глагола в этой модели. Можно пред-

положить, что оставшиеся случаи, не поддающиеся классификации по указанным критериям, относятся к области фразеологии. Если такое предположение верно, то само собою напрашивается вывод, что изоморфизм плана выражения и плана содержания кончается там, где начинается идиоматика. Однако делать какие-либо определенные выводы по этому вопросу еще рано, так как он может и должен служить темой самостоятельного исследования.

Выше уже говорилось, что для систематизированного описания значений того или другого слова необходимо установить их иерархию в зависимости от иерархии выражающих их форм. Критериями для определения иерархии языковых единиц, как уже говорилось, можно считать широту сферы употребления единицы, степень сложности ее структуры и способность ее к словообразованию.

По этим критериям мы установили первичность модели № 1 и детерминируемого ею значения.

Можно ли при помощи этих же критериев установить иерархию значений внутри одной и той же модели? Легко убедиться, что первый и третий из них остались в силе.

Попытаемся на основе широты сферы употребления и способности к словообразованию каждой из выделенных подгрупп, оформляющих какое-нибудь специализированное значение *make*, установить иерархию этих подгрупп и соответствующих значений.

На первый взгляд может показаться, что наиболее широкая сфера употребления характеризует сочетания, принадлежащие к группе II. Статистические данные показывают, что доля этой группы составляет примерно 52% всех случаев употребления *make* в модели № 1. Однако не следует забывать, что иерархия значений относится к языку, а не к речи, тогда как данная статистика отображает именно речь, функционирование системы языка в речи.

Для установления иерархии значений существенным должно являться определение круга тех единиц, с которыми может сочетаться *make* в данном значении, частотность же употребления этих сочетаний в речи нерелевантна. Определение круга сочетаемости *make* в различных группах показывает, что наиболее широкой сферой употребления обладает IV группа (примерно 48% количества всех существительных, с которыми сочетается *make* в исследуемом материале). Значит, по первому критерию первичным должно считаться значение, детерминируемое группой IV. Это подтверждается также и тем, что именно эта группа образует большинство производных от *make* (до 75%), о которых мы говорили при установлении первичности моделей № 1. Традиционная лексикография определяет это значение *make* как творческую созидательную деятельность, направленную на изготовление, создание чего-то.

Исходя из всего изложенного, можно сделать следующие выводы.

1. Структурные методы, в частности методы дистрибутивного и трансформационного анализа, обеспечивают объективные критерии для разграничения значений многозначного слова.

2. Надежными критериями для разграничения значений многозначного слова можно считать:

а) выделенные путем дистрибутивного и трансформационного анализа текстов конструктивные модели, характерные для данного слова;

б) совокупность характерных для данного слова дистрибуций, подвергающихся данной трансформации или совокупности трансформаций, внутри каждой из выделенных моделей.

3. Каждое отдельное значение многозначного слова является функцией характерных для него дистрибуций и трансформаций.

4. Применение методов дистрибутивного и трансформационного анализа и метода непосредственно составляющих для дифференциации зна-

чений многозначного слова возможно до определенного предела. Можно предположить, что материал, не поддающийся исследованию указанными методами, относится к области фразеологии.

5. Критериями для установления иерархии языковых единиц и оформляемых ими значений можно считать: широту сферы употребления данной единицы, степень сложности ее структуры и способность ее к словообразованию.

6. На основании этих критериев устанавливаем, что для *make* первична модель № 1 и детерминируемое ею значение, а внутри этой модели — группа IV, оформляющая значение, определяемое традиционной лексикографией как «творческая созидательная деятельность, направленная на изготовление, создание чего-нибудь».

ПРИЛОЖЕНИЕ

ИЛЛЮСТРАЦИЯ К СХЕМЕ РАЗЛИЧЕНИЯ ЗНАЧЕНИЙ В МОДЕЛИ № 1

1 — He'll make an announcer.
2 — We'll make a good double.
3 — Your Christianity'll make a bankrupt.
4 — 8 gallons make a bushel.
5 — Make a criticism.
6 — Make one's debut, m. comradship.
7 — Make amends.
8 — Make an addition, a correction, a bargain, a date.
9 — Make way.
10 — She made a convert. We made six new readers.
11 — Who made the turkey?
12 — Make an ally.
13 — We can't make a new Cabinet before Magnus comes in.
14 — Make an armistice, m. an anarchy.
15 — Make business, m. money, m. capital.
16 — Make a car, a boiler, an alloy, brass, coke. . .
17 — Make an angle.
18 — Avarice, lechery, sloth, jealousy, pride made the crusades.
19 — Shall we make the station by nine?

ON STRUCTURAL METHODS IN LEXICOLOGY

Summary

The article is a short account of an attempt to employ distributional and transformational methods as well as the method of immediate constituents for discriminating the meanings of the verb *make* known as one of the highly polysemantic verbs in the English language.

The analysis suggests the following conclusion: structural methods in differentiating the specific meanings of the verb are effective within definite limits. The remaining material (7 p. c.) which defies analysis may be supposed to constitute the realm of phraseology.

П. А. Соболева

КОМПОНЕНТНЫЙ АНАЛИЗ ЗНАЧЕНИЙ ГЛАГОЛА НА ОСНОВЕ СЛОВООБРАЗОВАТЕЛЬНОГО ПРИЗНАКА

Компонентный анализ значений преследует двойную цель:

1) Организацию определенной части семантической субстанции языка, т. е. распределение значений по классам, подклассам, группам, подгруппам и т. д., находящимся в определенной иерархической зависимости друг от друга, или, иными словами, распределение значений по семантическим полям [1] разного порядка.

2) Определение лингвистических значений, репрезентируемых этими семантическими полями, а также выявление идеальных отношений между ними, т. е. структуры реляционного языкового каркаса[2].

В основу компонентного анализа может быть положен любой структурный признак исследуемой категории слов: синтаксический[3], морфологический или словообразовательный.

В настоящем исследовании компонентный анализ глагольных значений проводится на основе дистрибутивных свойств производных отглагольных существительных типа *run*, *break*, *wish* [4], т. е. способности этих существительных выступать в обобщенных значениях, имеющих определенное структурное выражение.

Метод настоящего исследования — дедуктивный, т. е. исследование распадается на серию анализов, в которой результат предыдущего является исходным моментом для последующего [5].

Каждый анализ заключается в определении линии сечения семантического пространства на два компонента. Для этого семантическое пространство (семантическая субстанция) рассекается произвольно до тех пор, пока не будет обнаружено правильное, т. е. структурно оправданное сечение (trial and error method) [6].

Правильность сечения определяется диагностирующим фактором. В качестве диагностирующего фактора выступают дистрибутивные свойства отглагольных существительных:

---

[1] См. S. Ullmann. The Principles of Semantics. Glasgow, 1951. S. Öhman. Theories of the Linguistic Field. «Word», Vol. IX, № 2, 1953.

[2] См. С. К. Шаумян. Структурные методы в семантике. «Лексикографический сборник», № 5 (в печати).

[3] См. Ю. Д. Апресян. О понятиях и методах структурной лексикологии (см. настоящий сборник, стр. 141—162).

[4] См.: П. А. Соболева. Словообразовательные отношения по конверсии между глаголом и отглагольным существительным в современном английском языке. Автореферат кандидатской диссертации, М., 1959.

[5] L. Hjelmslev and N. Y. Uldall. Outline of Glossemantics. Copenhagen, 1957.

[6] I. Siertsema. A Study of Glossemantics. The Hague, 1955.

а) Формула конструкции [7]. Если формула конструкции недостаточно индивидуальна, она уточняется соответствующим трансформационным типом [8].

б) Формула субституции.

Диагностирующая формула конструкции определяется следующим образом:

1. Все высказывания, содержащие элементы, относящиеся к предполагаемым компонентам, сводятся к элементарным бинарным конструкциям (или грамматическим конфигурациям [9]).

2. Отбираются только те конструкции, которые содержат исследуемые слова. Каждая конструкция отбирается только один раз. Отобранные конструкции кодируются [10] для удобства сравнения.

3. Закодированные конструкции сравниваются между собой и одинаковые отбрасываются.

4. Если остатка не получается, деление произведено ошибочно. Если в остатке получается минимум одна элементарная конструкция, по отношению к которой значения слов последовательно противопоставляются друг другу, то деление произведено правильно, и остаточная конструкция является диагностирующей.

Проиллюстрируем операции по нахождению диагностирующей конструкции на материале глаголов движения [11]. Предполагаемые компоненты — «прерывистые» и «непрерывные» движения.

1. Сведение к элементарным бинарным конструкциям. Отбор и кодирование конструкций для последующего сравнения

А. Для постулируемого семантического поля «непрерывных движений».

Высказывание I: *He would like to get a ride* → *get a ride* → V Nv.

Высказывание II: *We all know you are perfectly happy slopping abou. the Manderly gardens and never breaking out of a slow walk* → *break out of a walk* → V *out of* Nv; *slow walk* → A Nv.

Высказывание III: *Dave broke into a run* → *break into a run* → V *into* Nv.

Высказывание IV: *When she got back into the drawing room the others were on the move* → *be on the move* → V *on* Nv.

Высказывание IV: *The news did something to arrest the run of the new doctor's unpopularity* → *the run of unpopularity* → Nv *of* N.

Высказывание V: *There's the drive back and I have to work to morrow* → → *the drive back* → Nv D.

Высказывание VI: *I'll go for another stroll* → *go for a stroll* → V *for* Nv.

---

[7] Р. С. Гинзбург и С. С. Хидекель. О работе над свободным словосочетанием. «Иностранные языки в школе», 1958, № 3.

[8] D. S. Worth. Transform Analysis of Russian Instrumental Constructions. «Word», vol. XIV, 1958; Z. S. Harris. Co-occurrence and Transformation in Linguistic Structure, «Language», vol. XXXIII, № 3, 1957.

[9] Т. Н. Молошная. Опыт статистического обследования грамматических конфигураций в английском математическом тексте. «Машинный перевод и прикладная лингвистика». М., 1960, № 4.

[10] При кодировании используется следующая система условных обозначений: N — существительное и его функциональные эквиваленты; Nv — отглагольное существительное; V — глагол и его функциональные эквиваленты; A — прилагательное и его функциональные эквиваленты; D — наречие и его эквиваленты, в частности сочетание предлога с существительным, заменяющееся наречием, например: *a drive to the station*, *a jump from the chair*, функционирующие так же, как *a drive back*, *a jump aside*; P — предлог. В тех случаях, где предлоги, артикли и глаголы *be* и *have* имеют существенное для наших целей структурное значение, они не кодируются. Если артикль не имеет дифференциального значения, он при кодировании опускается.

[11] О том, как выделяются сами глаголы движения, см. ниже, стр. 184.

Остальные высказывания сводятся к тем же элементарным конструкциям либо непосредственно, либо после некоторых операций по сворачиванию [12] их членов. Следовательно, операция редукции и подготовки конструкций для сравнения окончена.

Б. Для постулируемого семантического поля «прерывистых движений».

Высказывание I: *He walks quickly and once again has to make a sudden jump aside when a lorry screeches at him → to make a jump →* V Nv; *sudden jump →* A Nv: *a jump aside →* Nv D.

Высказывание II: *When the shape stilled and flattened itself for a spring, George lifted his rifle and fired → flatten oneself for a spring →* V N *for* Nv. V N сворачивается в V[13], следовательно получаем: V *for* Nv.

Высказывание III: *She was an expert at tennis, as she later proved, putting Gregory in a heavy perspiration and keeping him on the jump → keep him on the jump →* V N *on* Nv. Сворачиваем V N в V и получаем: V *on* Nv.

Высказывание IV: *He began to show them furniture, padding up and down the aisles of the emporium with a stately tread → pad up and down the aisles with a stately tread. Up and down the aisles* можно выразить как P N, заменяющееся D, а V D сворачивается в V [14]. Следовательно, окончательная формула принимает вид: V *with* Nv (*stately tread* сворачивается в Nv).

Высказывание V: *He indicated the secretaries with the jerk of his hand→jerk of his hand →* Nv *of* N.

Все новые высказывания сводятся к тем же элементарным конструкциям, следовательно операция редукции закончена и для второго предполагаемого компонента.

2. Операция сравнения элементарных конструкций

«Непрерывные движения»	«Прерывистые движения»
V Nv	V Nv
V *out of* Nv	—
V *into* Nv	—
A Nv	A Nv
V *on* Nv	V *on* Nv
Nv *of* N	Nv *of* N
Nv D	Nv D
V *for* Nv	V *for* Nv
—	V *with* Nv

После отбрасывания совпадающих конструкций в остатке получаем V *into/out of* Nv, с одной стороны, и V *with* Nv, с другой. Очевидно, разложение на компоненты проведено правильно. Конструкция V *into/out of* Nv диагностирует некоторое более маркированное (специализированное), нежели общее значение движения, лингвистическое значение, которое удобно назвать «непрерывное движение», а конструкция V *with* Nv диагностирует другое маркированное значение, которое можно назвать «прерывистым движением». Следовательно, исходные значения глаголов движения вполне правомерно разложить на два компонента: «прерывистых» и «непрерывных движений».

Диагностирующая формула субституции определяется следующим образом. Выбирается достаточно широкая формула конструкции, на

---

[12] Z. S. Harris. From Morpheme to Utterance. «Language», vol. XXII, № 3, 1946; R. S. Wells. Immediate Constituents. «Language», vol. XXIV, № 1, 1948.

[13] Переходный глагол с объектом имеет такую же дистрибуцию как непереходный глагол (по отношению к субъекту и адвербиальному члену).

[14] Глагол с адвербиальным членом имеет дистрибуцию, равную дистрибуции глагола без такового.

которую положительно реагируют слова обоих предполагаемых компонентов, т. е. какая-либо конструкция X Nv, где в качестве Nv выступают как производные от одних, так и производные от других глаголов.

Затем на место X подставляются индивидуальные члены класса X до тех пор, пока не будет обнаружено такое слово, которое удовлетворяет одним и не удовлетворяет другим исследуемым существительным. Если такое слово будет найдено, т. е. исследуемые элементы будут последовательно противопоставляться друг другу по найденному члену класса X, то диагностирующая формула субституции определена правильно.

Таким образом, формула конструкции выражается в терминах классов слов и строевых элементов языка [15], а формула субституции выражается как в терминах классов слов, так и в терминах индивидуальных слов.

Проиллюстрируем операции по нахождению диагностирующей формулы субституции на том же материале глаголов движения. Возьмем одну из общих формул конструкций, отброшенных нами во время предыдущих операций, например конструкцию V Nv, на которую положительно реагируют как «непрерывные», так и «прерывистые движения». Посмотрим, какие глаголы можно подставить на место V. Ряд испытаний показывает, что глагол *give* может употребляться только с существительными, производными от «прерывистых» процессов: *to give a jump, to give a jerk, to give a start, to give a bounce, to give a stagger.*

Следовательно, формула *to give* Nv диагностирует значение прерывистости движения.

С другой стороны, в такой общей для обоих компонентов формуле, как V *for* Nv, могут употребляться глаголы *go, start, set off* только с существительными, производными от «непрерывных» процессов движения, например: *go for a swim, start for a run, set off for a walk, go for a sail, go for a drive, go for a ride, go for a hike, go for a crawl* и т. д. Для наших целей достаточным является один из членов обнаруженной группы, например глагол *go*. Формула субституции *go for* Nv диагностирует непрерывность движения исходного глагола.

Итак, в результате проведенных операций мы обнаружили две формулы конструкции и две формулы субституции, диагностирующие некоторые более маркированные значения, нежели исходное значение процесса движения.

В дальнейшем при изложении результатов компонентного анализа мы будем пользоваться полученными нами диагностирующими формулами, не останавливаясь на операциях по их получению.

Следует отметить, что три из полученных выше четырех диагностирующих формул являются избыточными. Необходимой и достаточной информацией о наличии двух компонентов является одна диагностирующая формула (конструкции или субституции).

Переходим к последовательной сегментации семантического пространства с помощью диагностирующих факторов.

3. Семантические поля переходных и непереходных значений

Значения глаголов прежде всего последовательно противопоставляются с точки зрения отношения обозначаемого ими действия к объекту.

Отношение действия к объекту является коррелятором [16], который в данном случае связывает семантическую субстанцию с реляционным каркасом языка. Обозначим его через α. По отношению к коррелятору α значения глаголов распределяются по двум семантическим полям первого

[15] Ch. C. F r i e s. The Structure of English. New York, 1952.

[16] С. К. Ш а у м я н. Структурализм как имманентная теория языка. М., 1958.

порядка со значениями переходности и непереходности. Обозначим их соответственно через a_t и a_i. Выделение семантических полей, образованных переходными и непереходными значениями, имеет достаточное структурное обоснование и без обращения к производным глагола[17] (наличие объекта и соотношение активной и пассивной трансформации [18]). Однако дистрибутивные свойства отглагольных существительных могут быть использованы для дальнейшего членения этих семантических полей первого порядка.

Так, представляется возможным разложить семантическое поле переходных значений на два компонента с точки зрения их отношения к трансформации «исходный глагол+объект». Если формула трансформации: V N → [19] N *has* Nv, то исходный глагол можно отнести к семантическому полю второго порядка с лингвистическим значением «воздействие на объект, вызывающее его качественное изменение», например:

to tear a curtain → the curtain has a tear
to cut a finger → the finger has a cut
to scratch a mirror → the mirror has a scratch
to break a wall → the wall has a break
to stain the cloth → the cloth has a stain
to polish one's manners → one's manners have a polish
to crease paper → the paper has a crease

Диагностирующая конструкция N *has* Nv в данном случае — показатель обобщенного значения «результат действия», в котором выступают отглагольные существительные, производные от глаголов со значением «воздействие на объект, вызывающее его качественное изменение». Другие глаголы не отвечают такой трансформации:

to drink water ←↦ [20] **water has a drink*
to spoil a table-cloth ←↦ **the table-cloth has a spoil*

Назовем второй компонент, отрицательно реагирующий на формулу N *has* Nv, полем второго порядка с лингвистическим значением «действие, не вызывающее качественного изменения объекта». Обозначим лингвистические значения обоих полей второго порядка соответственно через a и b.

Второй компонент может быть подвергнут дальнейшей сегментации в зависимости от отношения к диагностирующей формуле N *is* Nv, где N является объектом или субъектом действия исходного глагола V. Например: *He replies «yes» ←→ «yes» is a reply. He bores people ←→ He is a bore.* Объединим глагольные значения этого типа в семантическое поле третьего порядка и назовем его лингвистическое значение «субъектно-объектным» (производные существительные выступают со значением субъекта или объекта действия). Обозначим его через b_1.

Глаголы противоположного поля не дают такой трансформации:

She mends a dress ←↦ **She (the dress) is a mend.*
He addresses the crowd ←↦ **He (the crowd) is an address.*

---

[17] Хотя структурное различие между этими двумя категориями значений находит отражение и в области производных. Так Фаулер (Fowler) в словаре «A Dictionary of Modern English Usage» пишет: «The suffix *-able* may be appended to any transitive verb to make an adjective with the sense. . . *-liable to be. . . -ed.*» Следовательно, способность глагола присоединять суффикс *-able* может рассматриваться как своего рода дифференциальный признак семантического поля переходных значений. Отрицательная реакция на формулу V *-able* указывает на принадлежность значения глагола к противоположному полю «непереходных».

[18] N. Chomsky. Syntactic Structures. 's-Gravenhage, 1957; R. B. Lees. The Grammar of English Nominalization. «International Journal of American Linguistics», vol. XXVI, № 3, 1960.

[19] Знак → трансформация в одном направлении, знак ←→ в обоих направлениях.

[20] Значок ←↦ — «невозможная трансформация».

Назовем значение этого поля условно «несубъектно-объектным» и обозначим b_2.

Семантическое поле «субъектно-объектных» значений представляется возможным разложить на два компонента в зависимости от исходного трансформа диагностирующей формулы N *is* Nv. Если формула трансформации VN, то исходные глаголы можно объединить в семантическое поле четвертого порядка с лингвистическим значением «объектно-переходные глаголы», например:

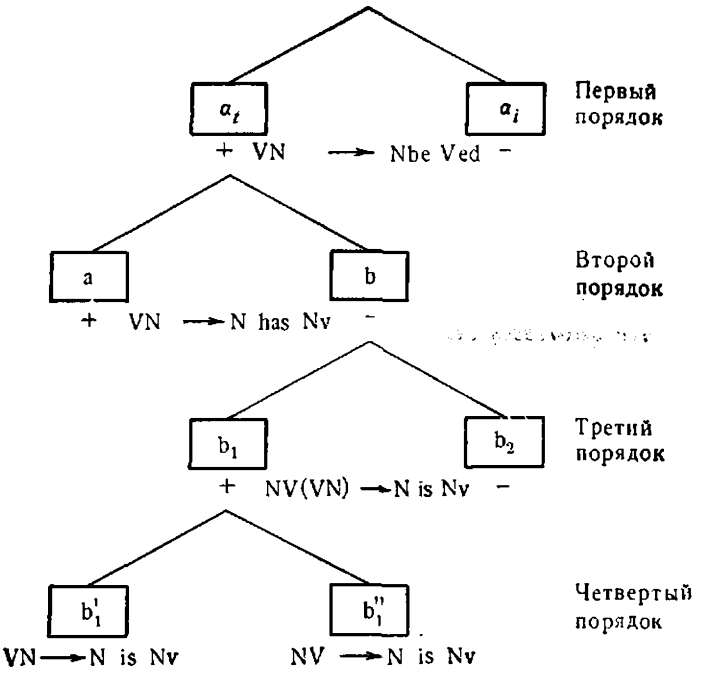
to grant land → the land is a grant
to catch fish → the fish is a catch
to win $ 20 → $ 20 is a win
to drink soda-water → soda-water is a drink
to envy good fortune → good fortune is one's envy
to love a girl → the girl is one's love.

Если трансформом N *is* Nv является N V, то исходные глаголы можно назвать «субъектно-переходными», например:

He tramps (the country) → He is a tramp.
The shawl wraps (one's shoulders) → The shawl is a wrap.
The sight delights (us) → The sight is a delight.
John comforts (his mother) → John is a comfort. . .
The sound alarms (us) → The sound is an alarm.
He disgraces (his family) → He is a disgrace. . .
Cold menaces plants → Cold is a menace. . .

Обозначим лингвистические значения полей четвертого порядка через b'_1 и b''_1.

Получив семантические поля четвертого порядка, считаем компонентный анализ по коррелятору а законченным, так как полей высшего порядка по принципу различных отношений действий к объекту на нашем материале пока не удалось обнаружить.



4. Семантические поля процессов и состояний

Вернемся к исходной экспериментальной базе и посмотрим, можно ли подвергнуть семантическое пространство членению по какому-либо другому принципу (коррелятору).

Анализ показывает, что традиционное деление значений глаголов на «процессы» и «состояния» имеет структурное обоснование. Так, отглагольные существительные, производные от глаголов состояний, положительно реагируют на формулу V *in* (*to*) Nv, например: *to lie in wait, to keep one's face in repose, to fall into disgrace, to be in bloom, to go off/into a swoon, to lie in sleep, to fall into a slumber, to look at somebody in scorn, to part in disgust, to fall in love, to be in trouble, to be in despair, to stop in wonder, to stare at somebody in fresh concern, to live in disgrace, to fly into a rage, to observe something in regret, to live in comfort* и т. д.

Эти же существительные положительно реагируют на формулу субституции *state of* Nv. Например:

state of { *disgrace, concern, wonder, fear,*
sleep, wait, repose, regret, trouble,
rage, comfort, love и т. д.

Существительные, производные от глаголов со значением процессов, отрицательно реагируют на последнюю формулу: **state of run*, **state of break*. Отношение к формуле *state of* Nv является крайним случаем противопоставления процессов и состояний. В формуле V *in* (*to*) Nv эти противоположности как бы нейтрализуются, ср. *break into a run* и *fly into a rage*. Однако полного снятия противоположности не происходит, что следует из возможности замены *into* на *out of* в случае процессов [21] и невозможности такой замены в случае состояния. Например, *break out of a slow walk*, но не **fall out of despair* [22].

Следовательно, формулы V *in* (*to*) Nv и *state of* Nv можно рассматривать как диагностирующие плоскость сечения между двумя полями первого порядка. Положительная реакция на эти формулы сигнализирует значение «состояний» (β_s), отрицательная реакция — значение «процессов» (β_p).

Семантическое поле с лингвистическим значением «состояния» представляется возможным разложить на два поля второго порядка в зависимости от отношений к диагностирующей формуле субституции *to feel* Nv.

Положительная реакция на эту формулу отличает исходные глаголы «моральных состояний» от глаголов «физических состояний». Обозначим данные лингвистические значения соответственно через *c* и *d*.

К полю со значением *c* относятся такие глаголы, как *alarm, dismay, confort, concern, respect, wonder, desire, hate, surprise, rage, scorn, love, need, want, envy, distress, favour, pity, hope, mistrust* и т. д.

Ко второму полю — такие, как *hurry, sleep, fuss, fret, repose, slumber, swoon, bloom, whirl, flutter, tremour, disarray, ruin, wreck, mess, blend, huddle* и т. д.

Компонент со значением *c* представляется возможным разложить на два семантических поля третьего порядка в зависимости от отношения к формуле N_1'*s* Nv *for* N_2, являющейся трансформом конструкции N_1 V N_2, например:

John loves Mary → John's love for Mary...
John pities Mary → John's pity for Mary...
John respects Mary → John's respect for Mary...
John mistrusts Mary → John's mistrust for Mary...
John regrets the past → John's regret for the past...

---

[21] См. стр. 177.

[22] Употребление типа *fall out of love* подразумевает определенный стилистический эффект.

Другие глаголы, производные существительные от которых положительно реагируют на формулу *to feel* Nv, имеют другие трансформации:

John disgusts Mary → Mary's disgust for John...
His health alarms Mary → Mary's alarm for his health...,

т. е. $N_1VN_2 \rightarrow N_2$*'s* Nv *for* N_1, а также другие трансформации, отличные от диагностирующей N_1 V $N_2 \rightarrow N_1$*'s* Nv *for* N_2.

Объединим первые в семантическое поле третьего порядка с лингвистическим значением «испытывание моральных состояний», а вторые — в противоположное однопорядковое поле со значением «приведение к моральным состояниям». Обозначим их условно как c_1 и c_2.

Семантическое поле третьего порядка со значением «испытывание моральных состояний» представляется возможным разложить еще на два компонента в зависимости от теста на конструкцию Nv *to* V (*desire to read*). Существительные, производные от таких глаголов, как *respect, regard, pity, concern, regret, love, scorn, relish, wonder, trust, hate, mistrust* и т. д., отрицательно реагируют на эту конструкцию, а такие, как *resolve, desire, want, wish, need, dislike, hope* и т. д., — положительно.

Первые можно назвать глаголами со значением «испытывание чувств», а вторые — «испытывание побуждений». Обозначим их c_1' и c_1''. Эти значения соответствуют семантическим полям четвертого порядка.

Вернемся к семантическому полю второго порядка со значением «физические состояния» (d). Это поле можно, в свою очередь, разложить на два поля третьего порядка с лингвистическими значениями «испытывание физических состояний» (d_1): *hurry, sleep, fuss, fret, repose, slumber, swoon, bloom, whirl, swirl* и т. д. и «приведение к физическим состояниям» (d_2): *disarray, smash, wreck, ruin, blend, mess, huddle* и т. д. Производные от них существительные различаются отношением к трансформации *be* V *-ed* → V *in* (*to*) Nv, например:

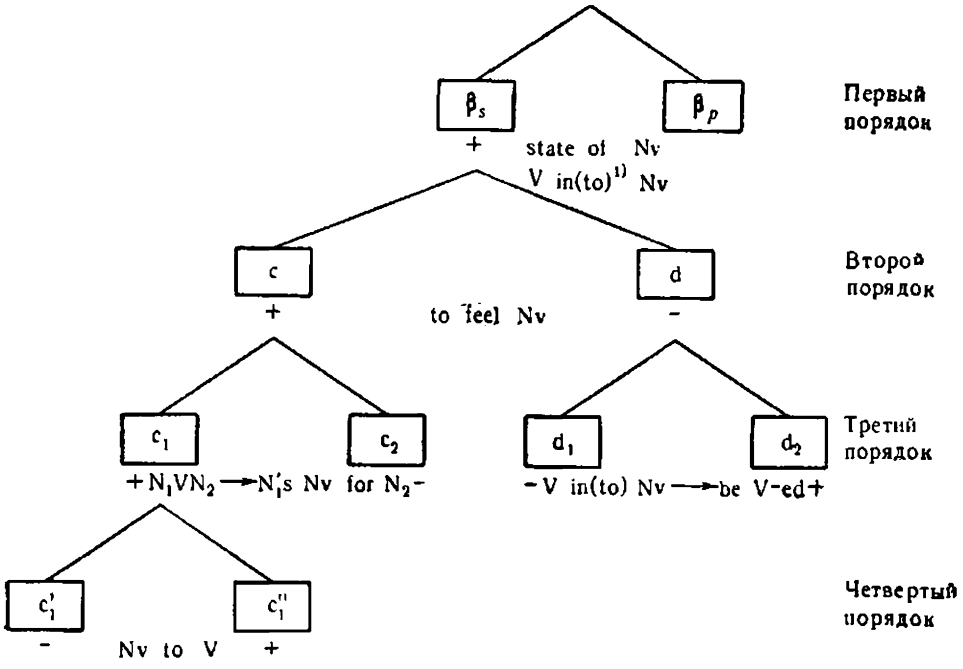
$$\textit{to be} \left\{ \begin{array}{l} \textit{ruined} \\ \textit{wrecked} \\ \textit{disarrayed} \\ \textit{messed} \\ \textit{blended} \\ \textit{huddled} \end{array} \right\} \rightarrow \textit{to be in} \left\{ \begin{array}{l} \textit{ruin} \\ \textit{a wreck} \\ \textit{disarray} \\ \textit{a mess} \\ \textit{a blend} \\ \textit{a huddle} \end{array} \right\}$$

Остальные существительные не дают такой трансформации: *to be in a fuss, to be in bloom, to be in tremour* и т. д. ↚ **to be fussed*, **to be bloomed* и т. д.

Полей более высоких порядков на материале глаголов состояний нам не удалось обнаружить.

Рассмотрим поле первого порядка со значением процессов. Анализ показывает, что обнаруженное выше противопоставление прерывистых и непрерывных движений распространяется на процессы вообще (в отличие от состояний). Последнее подтверждается проверкой на диагностирующие формулы V *with* Nv: *to close with a bang, to open with a crack, to meet with a sob, to fall with a thump, to answer with a smile, to fall with a burst, to cut with a slash, to swing with a click, to shut with a snap, to break with a roar, to wink with a laugh* и т. д., и на формулу *to give a* Nv [23]: *to give a bite, to give a slash, to give a crack, to give a laugh, to give a sob, to give a click, to give a smile, to give a wink, to give a glance* и т. д.

[23] Проверка формулы *give* Nv на более широком материале потребовала ее уточнения — включения неопределенного артикля в ее состав как обязательного элемента. Именно артикль отличает конструкции типа *give a bite* от конструкций типа *give trouble, give delight*, в которых могут употребляться существительные, производные от глаголов состояния.



В этих конструкциях отглагольное существительное выступает в обобщенном значении акта действия.

Существительные, производные от глаголов, обозначающих непрерывные процессы, в таких конструкциях не употребляются (непрерывный процесс не распадается на акты).

Конструкцию V *with* Nv, диагностирующую лингвистическое значение прерывистости процесса, не следует смешивать с конструкцией, внешне ее напоминающей: *to treat somebody with respect, to look with alarm, to regard somebody with concern, to accept something with trust* и т. д., в которой употребляются существительные, производные от глаголов состояний. Различие между этими конструкциями вскрывается при сравнении их трансформов. Существительные, производные от глаголов состояний, не трансформируются в причастие I при устранении предлога *with*, а трансформируются только в наречие с суффиксом *-ly*:

to treat somebody with respect ↔ to treat somebody rescpetfully
to look at somebody with alarm ↔ to look at somebody alarmingly
to regard somebody with concern ↔ to regard somebody concernedly
to accept somethiny with trust ↔ to accept somethiny trustfully

Существительные, производные от глаголов со значениями прерывистых процессов, могут при устранении предлога *with* трансформироваться в причастие I:

to fall with a burst ↔ to fall bursting
to close with a bang ↔ to close banging
to break with a roar ↔ to break roaring
to meet with a sob ↔ to meet sobbing
to leave with hurried steps ↔ to leave stepping hurriedly

В отдельных случаях возможна и двойная трансформация:
to answer with a laugh ↔ to answer laughing → to answer laughingly;
to look with a smile ↔ to look smiling → to look smilingly.

Отношение к трансформу V_1 V-*ing* и отличает конструкцию типа V_1*with* Nv, диагностирующую прерывистость процесса, от словосочетаний типа *to treat somebody with respect.*

Итак, с точки зрения реакции исследуемых существительных на диагностирующие формулы *to give a* Nv, V *with* Nv ↔ V V-*ing* [24] исходные глаголы распределяются по двум семантическим полям второго порядка с лингвистическими значениями прерывистости и непрерывности процесса. Обозначим эти значения как *e* и *f*.

С другой стороны, исходное поле первого порядка с лингвистическим значением β_p можно разложить на два другие компонента в зависимости от отношения производных слов в формуле Nv P N ↔ Nv D, или, вернее, от их отношения к элементу P в этой формуле. Если в формуле Nv P N в качестве P выступают предлоги направления — *to*, *from*, *into*, *through*, то Nv является производным от глаголов движения: *approach to success*, *fall through space*, *rise from the chair*, *falls into pit-bogs*, *dips into villages*, *return into the kitchen*, *drive from the station*, *jump to the left*, *advance to civilization* и т. д.

Эти существительные могут употребляться и с другими предлогами, в частности предлогами места: *drive on the wagon*, *walk in the park*, однако во всех этих случаях их можно заменить диагностирующими предлогами направления: *drive to the station*, *walk from the park* и т. д.

Если же в конструкции Nv P N ↔ Nv D в качестве P выступают другие предлоги, кроме предлогов направления, то Nv является производным от глаголов действий: *work at a plant*, *crease on his trousers*, *design on the curtain*, *fight in the desert* и т. д. В данном случае замена предлогом направления невозможна.

Указание на трансформ Nv D важно для того, чтобы отличить данную конструкцию от такой, как, например, *give a pinch to a dog*, где тоже могут употребляться предлоги направления, но где сочетание предлога с существительным не заменяется наречием.

Перепишем диагностирующую формулу в следующем виде: Nv *to*/*from* N ↔ Nv D, выбрав в качестве представителей предлогов направления предлоги *to* и *from*.

Итак, тест на формулу Nv *to*/*from* N выявляет два семантических поля второго порядка с лингвистическими значениями «движений» и «действий». Обозначим их через *g* и *h*.

Рассмотрим поле со значением «движений». Данное поле можно разложить на два компонента по отношению отглагольных существительных к формуле V *on* Nv.

Значительная часть отглагольных существительных употребляется в этой конструкции, например: *to be on the float*, *to be on the hop*, *to catch something on the fly*, *to be on the move*, *to be on the tramp*, *to keep somebody on the jump*, *to catch something on the bounce*, *to be on the prowl*, *to be on the march*, *to be on the turn*, *to be on a tour* и т. д.

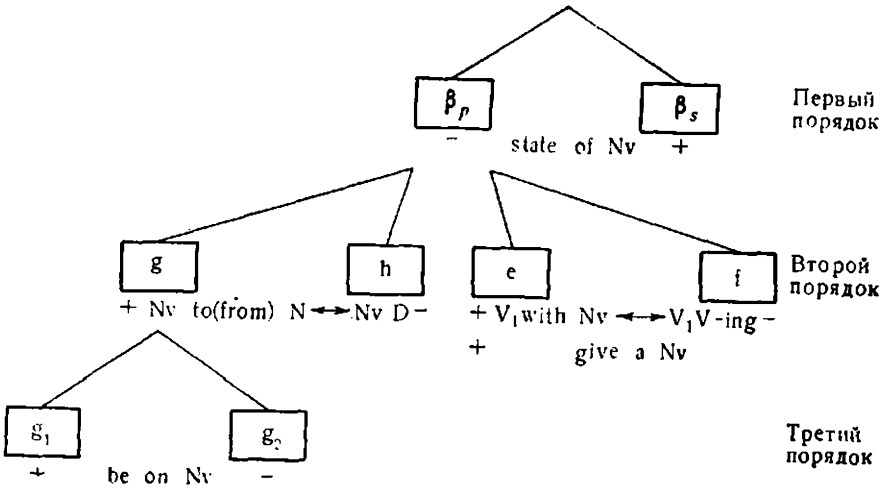
Другие существительные относятся к этой конструкции отрицательно, например такие как: *go*, *retreat*, *advance*, *plunge*, *waddle*, *amble*, *limp*, *scramble*, *waltz*, *dance* и т. д. Значение последних можно определить как «направленные и качественные перемещения», а первых как «просто перемещения». Обозначим эти значения через g_1 и g_2. Значения g_1 и g_2 соответствуют семантическим полям третьего порядка.

Аналогичному компонентному анализу может быть подвергнуто и семантическое поле второго порядка со значением действий.

Конечным этапом компонентного анализа должны явиться далее не сводимые синонимо-антонимические группы значений.

---

[24] Формулы V *into*/*out of* Nv и *go for* Nv, выведенные нами на стр. 177 и 178, оказались действительными только для глаголов движения.



5. Границы семантических полей и соотношение их дифференциальных признаков

Как следует из нашего изложения, семантические поля выделяются по одному или нескольким диагностирующим факторам, которые можно назвать дифференциальными признаками семантических полей.

Естественно возникает вопрос, как соотносятся между собой эти признаки в тех случаях, когда их более одного. Например, семантическое поле «состояний» выделяется нами по двум признакам: по формуле конструкции V *into* Nv и по формуле субституции *state of* Nv. Значит ли это, что все существительные, прозводные от этих глаголов, в равной степени положительно реагируют на обе формулы? На этот вопрос приходится ответить отрицательно. Существует некоторое количество глаголов, отрицательно реагирующих либо на первую, либо на вторую формулу. Например, существительные *regard* и *favour* не употребляются в формуле *state of* Nv, употребляясь при этом в формуле V *in* (*to*) Nv: *hold somebody in regard, speak in somebody' favour*. Существительные *delay, resolve, menace*, наоборот, употребляются в первой формуле: *state of delay, resolve, menace* и не употребляются во второй.

При выделении дополнительных дифференциальных признаков, сигнализирующих более маркированные значения, могут добавиться один или два элемента, выходящие за пределы семантических полей предыдущих порядков. Например, в семантическое поле четвертого порядка «испытывания моральных побуждений», отвечающее всем дифференциальным признакам низших порядков, то есть V *in* (*to*) Nv, *state of* Nv, *to feel* Nv, N_1 V $N_2 \rightarrow$ N_1*'s* Nv *for* N_2, Nv *to* V, включается глагол *wish*, отвечающий только трем последним формулам: *feel a wish, a wish for smth, a wish to do smth*, и не отвечающий первым двум: **state of wish*, **be in wish*.

Подобное соотношение признаков создает картину наложения и перекрещивания некоторых более элементарных полей внутри основного поля. Перекрещивание элементарных полей приводит к тому, что у основных полей получаются «расплывчатые» границы.

6. Области пересечения семантических полей

Естественно предположить, что в общем семантическом пространстве семантические поля, не связанные отношениями дизъюнкции, могут пересекаться друг с другом.

Для определения области сечения двух полей достаточно проверить все элементы одного поля на диагностирующую формулу другого поля.

Если при этом ни одно значение первого поля не отвечает положительно на проводимый тест, то рассматриваемые поля не имеют точек соприкосновения. Например, поле третьего порядка со значением «испытывание моральных состояний» c_1 (*respect, regard, pity, regret, scorn, desire, want, hate, love, resolve, need* и т. д.) не имеет области пересечения с полем второго порядка «воздействие на объект, вызывающее его качественное изменение», критерием отнесения к которому является положительная реакция на формулу VN → N *has* Nv (*to cut a finger → the finger has a cut*) (см. стр. 179). Или же поле третьего порядка «приведение к моральным состояниям» c_2 (*alarm, dismay, comfort, disgrace, distress, surprise, trouble, calm, menace* и т. д.) не пересекается с полем четвертого порядка «объектно-переходных» b_1', что показывает отрицательный тест на диагностирующую формулу V N → N *is* Nv (*to love Mary → Mary is one's love; to grant land → the land is a grant*) (см. стр. 180).

Если некоторая часть значений одного поля положительно реагирует на диагностирующую формулу другого поля, то данные поля имеют некоторую общую часть. Например, поле третьего порядка со значением «приведение к моральным состояниям» c_2 и поле четвертого порядка «субъектно-переходных» b_1''. Проверка на формулу N V ↔ N *is* Nv показывает, что поле «субъектно-переходных» значений отсекает часть поля — «приведение к моральным состояниям»:

N_1 *alarms* $N_2 \rightarrow N_1$ *is an alarm*
N_1 *comforts* $N_2 \rightarrow N_1$ *is a comfort. . .*
N_1 *disgraces* $N_2 \rightarrow N_1$ *is a disgrace. . .*
N_1 *lures* $N_2 \rightarrow N_1$ *is a lure. . .*
N_1 *surprises* $N_2 \rightarrow N_1$ *is a surprise. . .*
N_1 *troubles* $N_2 \rightarrow N_1$ *is a trouble. . .*
N_1 *dazes* $N_2 \rightarrow N_1$ *is a daze. . .*
N_1 *discomforts* $N_2 \rightarrow N_1$ *is a discomfort. . .*

Другие элементы этого поля показывают отрицательную реакцию на такую трансформацию:

N_1 *dismays* $N_2 \leftarrow\!\!\!\!\rightarrow {}^*N_1$ *is a dismay. . .*
N_1 *calms* $N_2 \leftarrow\!\!\!\!\rightarrow {}^*N_1$ *is a calm. . .*
N_1 *disdains* $N_2 \leftarrow\!\!\!\!\rightarrow {}^*N_1$ *is a disdain. . .*

Следовательно область пересечения проходит между первыми и вторыми значениями (см. рис. 1, тело kolqpr).

Поле третьего порядка «испытывание моральных состояний» c_1 обнаруживает некоторую общую часть с полем четвертого порядка «объектно-переходных», что выявляется проверкой на формулу VN → N *is* Nv:

N_1 *loves* $N_2 \rightarrow N_2$ *is* N_1*'s love*
N_1 *envies* $N_2 \rightarrow N_2$ *is* N_1*'s envy*
N_1 *relishes* $N_2 \rightarrow N_2$ *is a relish*
N_1 *wishes* $N_2 \rightarrow N_2$ *is* N_1*'s wish*
N_1 *needs* $N_2 \rightarrow N_2$ *is* N_1*'s need*
N_1 *desires* $N_2 \rightarrow N_2$ *is* N_1*'s desire*

(см. рис. 1, тело inmlok).

Другие глаголы отвечают отрицательно на данный тест и, следовательно, остаются за областью сечения с полем «объектно-переходные значения», например:

N_1 *respects* $N_2 \leftarrow\!\!\!\!\rightarrow {}^*N_2$ *is* N_1*'s respect*
N_1 *pities* $N_2 \leftarrow\!\!\!\!\rightarrow {}^*N_2$ *is* N_1*'s pity*

Если все элементы одного поля одинаково реагируют на диагностирующую формулу другого поля, то первое поле целиком входит в состав вто-

рого. Например, поле со значением «воздействие, вызывающее качественное изменение объекта» и поле «переходные значения». Или поле третьего

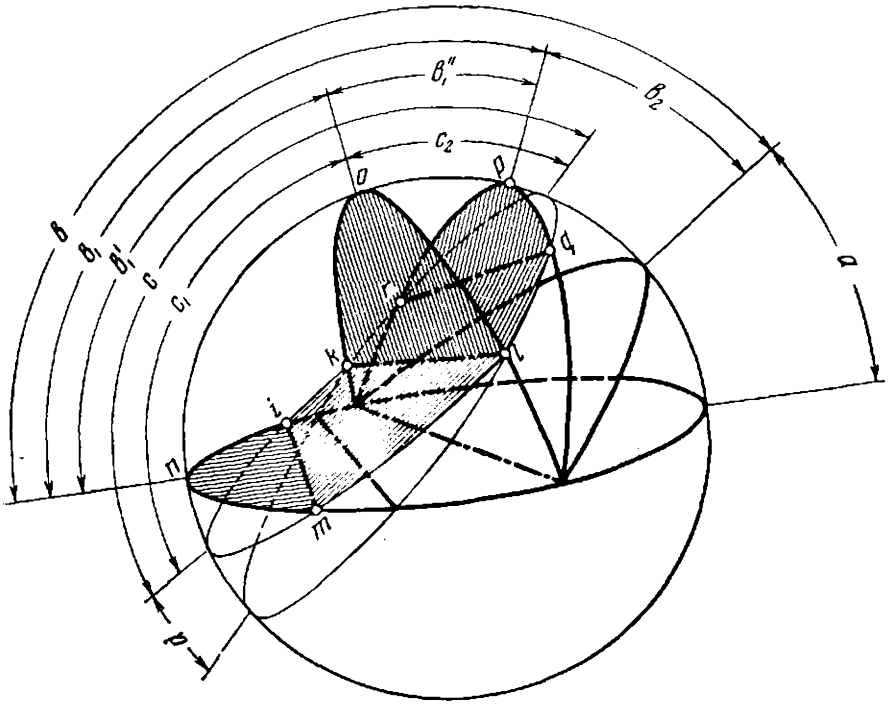


Рис. 1

порядка «испытывание физических состояний» (*sleep*, *fuss*, *fret*, *repose*, *slumber*, *whirl*, *flutter*, *swoon*, *bloom* и т. д.), целиком входящее в поле первого порядка «непереходных» значений, что следует из отрицательной реакции всех элементов на формулу *be* V *-ed*.

7. Функции лингвистических значений

Проведя компонентный анализ некоторой части семантической субстанции языка, можно ответить на вопрос о том, каковы взаимоотношения между обнаруженными лингвистическими значениями.

Так, однопорядковые значения, соответствующие непересекающимся полям, могут находиться в отношениях интердепенденции, например значение «моральные состояния» и значение «физические состояния», или «переходные» и «непереходные» значения. Пользуясь условными обозначениями для лингвистических значений, запишем функцию, в которую вступают эти величины как c V d; α_t V α_i, где V обозначает функцию дизъюнкции [25].

Лингвистические значения, соответствующие пересекающимся семантическим полям, могут находиться в отношении констелляции, т. е. не предполагать и не исключать друг друга. Например, значение «переходности» и значение «процесса» или значение «непрерывности процесса» и значение «перемещения». Эти отношения можно записать в виде функции конъюнкции: $\alpha_t \cdot \beta_p$; $f \cdot g_2$.

---

[25] Символическое обозначение функций: V, ., $\subset$, ϵ заимствовано нами из работы: S. K. Langer. Introduction to Symbolic Logic. New York, 1953.

Кроме этих функций, между разнопорядковыми полями могут устанавливаться отношения детерминации. Таковы, например, отношения между значением «состояние» и значением «моральное состояние» или между «переходными» значениями и значением «действие, вызывающее качественное изменение объекта». Запишем эти функции как: $c \in \beta_s$; $a \in \alpha_t$.

Поля, репрезентирующие значения c и a, целиком входят в соответствующие поля со значениями β_s и α_t в качестве их компонентов. При отношениях детерминации один член функции выступает в качестве постоянного члена, а другой переменного. В данных примерах постоянными являются β_s и α_t, а переменными c и a.

Разнопорядковые значения могут также находиться в отношениях импликации. Соответствующие им поля входят одно в другое, но не в качестве компонента, например поле со значением «испытывание физических состояний» и поле со значением «непереходность». Запишем эту функцию как $d_1 \subset \alpha_t$.

Таким образом, отношения между лингвистическими значениями, репрезентируемыми семантическими полями различных порядков, сводятся к четырем типам функций: функции дизъюнкции, функции конъюнкции, функции детерминации и функции импликации.

Определив лингвистические значения, репрезентируемые некоторой частью семантической субстанции языка, а также функции, в которые они могут вступать друг с другом, можно ответить на второй вопрос, поставленный в начале работы, а именно, какова внутренняя структура той части языкового каркаса, которая соответствует обнаруженным семантическим полям глагола. Внутренняя структура языкового каркаса обнаруживается через формулы конструктов [26], соответствующих обнаруженным семантическим полям глагола.

Под лингвистическим конструктом семантического поля следует понимать отношения, которые устанавливаются между его значением и другими лингвистическими значениями языка, т. е., иными словами, сумму функций, в которые вступает репрезентируемое этим полем лингвистическое значение.

Формула конструкта семантического поля должна указывать на место соответствующего ему лингвистического значения в общей системе лингвистических значений языка.

Воспользуемся обнаруженными нами лингвистическими значениями для выведения формул конструктов, соответствующих семантическим полям глагола.

$\alpha_t, \alpha_i, \beta_s, \beta_p$	I порядка
a, b, c, d, e, f, g, h	II порядка
$b_1, b_2, c_1, c_2, d_1, d_2, g_1, g_2$ . . .	III порядка
$b_1', b_1'', c_1' c_1''$	IV порядка

Какова, например, формула конструкта семантического поля III порядка с лингвистическим значением «приведение к моральным состояниям» c_2? Для выведения формулы конструкта этого поля надо учесть все те функции, в которые вступает репрезентируемое им значение.

С одной стороны, значение «приведение к моральным состояниям» c_2 детерминируется менее маркированным значением «моральное состояние» c, так как поле со значением c_2 является компонентом поля со значением c, т. е. $c_2 \in c$ (см. рис. 1).

---

[26] С. К. Шаумян. Структурализм как имманентная теория языка. М., 1958.

С другой стороны, значение c_2 исключает значение «испытывание моральных состояний» c_1, так как соответствующие им поля являются двумя компонентами поля предыдущего порядка (см. рис. 1.) Между значением c_2 и c_1 устанавливаются отношения дизъюнкции: $c_2 \vee c_1$.

Кроме того, значение c_2 влечет за собой значение «воздействие на объект, не вызывающее его качественного изменения» b, так как проверка на пересечение полей показывает, что поле со значением c_2 целиком включается в поле со значением b : $c_2 \subset b$ (см. рис. 1).

Помимо этого, значение c_2 предполагает значение «субъектная переходность» b_1'', так как соответствующие им поля при пересечении образуют некоторую общую часть (см. стр. 186 и рис. 1): $c_2 \cdot b_1''$.

Итак, значение c_2 вступает в функцию детерминации со значением c, в функцию дизъюнкции со значением c_1, в функцию импликации со значением b и в функцию конъюнкции со значением b_1''.

Сумма этих функций и определяет отношение c_2 к другим лингвистическим значениям языка, т. е. является конструктом репрезентирующего его семантического поля (сокращенно Кc_2):

$$\text{(I)}\quad \text{К}\,c_2 = \begin{cases} c_2 \in c \\ c_2 \vee c_1 \\ c_2 \subset b \\ c_2 \cdot b_1 \end{cases}$$

Данная формула примет более развернутый вид, если в нее включить и указание на место значений c, c_1, b и b_1'' в общей системе обнаруженных лингвистических значений:

$$\text{(II)}\quad \text{К}\,c_2 = \begin{cases} c_2 \in c \in \beta_s \vee \beta_p \\ c_2 \vee c_1 \in c \in \beta_s \vee \beta_p \\ c_2 \subset b \in \alpha_t \vee \alpha_i \\ c_2 \cdot b_1'' \in b_1 \in b \in \beta_t \vee \alpha_i \end{cases}$$

Формула (I) является краткой, а формула (II) — полной формулой конструкта семантического поля со значением c_2.

Аналогичным образом можно получить формулу конструкта для любого из обнаруженных нами семантических полей глагола.

Следует ожидать, что расширение базы исследования даст возможность построить модель многомерного семантического пространства со строгой иерархией семантических полей, место каждого из которых в общей системе определяется формулами соответствующих лингвистических конструктов.

COMPONENTAL ANALYSIS OF VERBAL MEANINGS ON THE BASIS OF WORD-BUILDING PROPERTIES

Summary

Componental analysis of verbal meanings has been employed in an attempt to arrange the semantic substance of a language into semantic fields of different orders.

In each set of operations two semantic fields are postulated, the resultant components being subjected to a similar procedure, etc. The formal criterion diagnosing a cross-section of the semantic substance is furnished by the distributional qualities of verbal nouns derived from the verbs analysed.

The resultant «semantic space» model presents a pattern of intersecting semantic gfields, whose hierarchial relations are expressed in terms of glossematics and symbolic logic.

Ю. К. Лекомцев

К ВОПРОСУ О СИСТЕМНОСТИ ГЛАГОЛОВ РЕЧИ В АНГЛИЙСКОМ ЯЗЫКЕ

Наша работа посвящена выявлению семантической модели для одной из лексических групп английского языка (группа глаголов речи). Выявление такой устойчивой семантической модели производилось путем сравнения нескольких синхронных состояний для данной лексической группы глаголов. Таким образом, историческое изучение определенной глагольной группы привело нас в данной работе к выводу о существовании синхронной семантической системы (модели), устойчивой относительно исторических изменений в языке. Следует подчеркнуть, что анализ семантических отношений, составляющих саму модель, не входит в задачу данной статьи.

Для каждого периода языка исследовалась небольшая группа наиболее распространенных глаголов речи, обладающих значениями «сказать», «говорить» [1].

Материалом работы послужили выборки из текстов древнеанглийского периода (эпический язык — язык прозы), среднеанглийского и новоанглийского периодов (список текстов см. в конце статьи).

Так как нам придется иметь дело с лексическими значениями, мы попытаемся использовать позиционную обусловленность значения в качестве фактора, контролирующего интуитивный анализ смысла.

Глагол, принадлежащий к определенной лексической группе, выступает в тексте связанным по значению с конкретными окружающими его словами. Так, глагол речи предполагает комплекс слов со значениями лица говорящего, лица слушающего, темы, объекта, вводимой речи (в форме прямой речи, косвенной речи и косвенного вопроса).

Под темой понимается косвенное дополнение при глаголе речи с предлогом *о, об* в русском языке, с предлогами *of, about* в английском языке. Под объектом понимается прямое дополнение при глаголе речи, имеющее либо значение единицы речи, как в др.-англ. *an ȝeþiode sprǣcon* 'говорили на одном языке' (букв.: 'говорили один язык'), либо служебное значение коррелята: *говорить все как было*; *сказать то, что не следует*', либо какое-нибудь иное значение. Различие между прямой и косвенной речью выражается в перемене лица при глаголе, в различии наклонения, в наличии союза при косвенной речи. В зависимости от индивидуального значения глагола наличествуют или преобладают статистически те или

[1] Перевод английских глаголов речи русскими глаголами, в частности глаголами *сказать* и *говорить*, весьма огрубленно передает смысл первых. Сопоставление с русскими *сказать* и *говорить* правомерно лишь в той степени, в какой для них и для соответственных английских глаголов речи является общим противопоставление по признакам: непротяженность и результативность действия — протяженность и нерезультативность действия.

иные синтаксические связи. Так, отсутствию объекта и лица слушающего у одного из глаголов речи и наличию объекта и прямой речи у другого будут соответствовать четко разграниченные значения [2]. Значение глагола, связанное с наиболее употребительными синтаксическими конструкциями при глаголе, будет считаться его основным, продуктивным значением [3]. Дадим краткую сводку материалов, дающую значение и конструкции глаголов речи для различных периодов языка.

КОММЕНТАРИИ К ТАБЛИЦЕ И ПРИМЕРЫ

I. Древнеанглийский период

а. Эпический язык.

Дополнительные данные по языку древнеанглийской [4] поэзии говорят о том, что *to* при лице слушающего встречается со всеми рассматриваемыми глаголами речи (и многими другими), за исключением глагола *secȝan*, при котором лицо слушающего стоит в дательном падеже без предлога. Для других памятников при *secȝan*, как и при *sprecan*, характерна тема с предлогом *ymb* (в «Беовульфе» употребляется с *fram*).

Примеры. *Maþelian.* Weard maþelode, ðær on wicȝe sæt, ombeht unforht: AEȝhwæþres sceal scearp scyldwiȝa ȝescād witan, worda ond worca, sē þe wēl þenceð [B. 286]: 'Стражник сказал, он сидел на коне, служитель храбрый: Каждый истинный воин, мыслящий здраво, должен разбираться в словах и поступках (делах)'. Wulfȝār maþelode tō his winedrihtne: «Hēr syndon ȝeferede, feorran cumene ofer ȝeofenes beȝanȝ Gēata lēode;» [B. 360]. 'Вулфгар сказал своему содружиннику: Сюда прибыли, переехав через простор океана, люди геатов' *Cwepan.* Cwædon þæt hē wǣre wyruldcyninȝa [B. 3180]. 'Сказали, что был он королем лучшим в мире' [говорили]

б. Язык прозы (IX—XI вв.).

Глагол *maþelian* отсутствует, его вариант *maþlān* малоупотребителен, то есть глагол фактически исчезает из языка. При глаголе *sprecan* лицо слушающего употребляется с предлогом *to* или *wiþ*. Глагол

---

[2] Например, в современном английском языке (материал взят из кн.: J. G a l s w o r t h y. The man of property) для глаголов *to say* и *to tell* основная масса случаев падает на употребление с объектом, с прямой (*to say*) и косвенной (*to say, to tell*) речью, т. е. подчеркивается то, что говорится. Для глаголов *to speak* и *to talk* большинство случаев приходится на непереходное употребление (вернее, на отсутствие как прямого, так и косвенного дополнения), т. е. подчеркивается самый процесс речи. С глаголом *to say* редко употребляется имя слушающего (5% из 200 случаев), что еще сильнее подчеркивает то, что говорится. С глаголом *to tell* имя слушающего употребляется часто (74% из 60 случаев), чем подчеркивается связь со слушающим, и т. д.

[3] Ср. следующие высказывания М. М. Покровского: «. . .Слова, сходные или прямо противоположные по значению, могут иметь свои собственные формальные особенности; далее, одно из них может даже передать некоторые из своих звуков другому; наконец, если слово, принадлежавшее к известной семасиологической категории, почему-либо выходит из нее, то оно может вместе с тем утратить и формальные особенности этой категории» (М. М. П о к р о в с к и й. Семасиологические исследования в области древних языков. «Избранные работы по языкознанию». М., Изд-во АН СССР, 1959, стр. 83). «В любом языке нам приходится чрезвычайно часто встречаться с переменой конструкции при одном и том же глаголе. Это зависит от того, что данный глагол в позднюю эпоху языка может ассоциироваться по значению с иными глаголами, чем в эпоху древнюю; понятно, что благодаря этой новой ассоциации глагол может потерять свою прежнюю конструкцию и приобрести конструкцию, которую имеют его новые соседи». . . (там же, стр. 85).

[4] См.: J. A l b e r s. Der syntaktische Gebrauch der Präposition *to* in der altenglischen Poesie. Kiel, 1907, стр. 17—18; J. J a c o b s o n. Der syntaktische Gebrauch der Präpositionen *for, ȝeon, of* und *ymb* in der ae. Poesie, Kiel, 1908, стр. 68.

	Древнеанглийский				
	Язык эпической поэзии (VIII в.)				
	Прямая речь	Косвенная речь	Объект	Тема	Имя слушающего
maþelian	+	—	—	—	[to] dat
cwæþan	(+)	+	+		
secȝan		+ þæt	+	fram, ymb	
talian					
sprecan	(+)		(+)*		to dat

	Среднеанглийский период (XIII—XIV вв.)				
	Прямая речь	Косвенная речь	Объект	Тема	Имя слушающего
quethen	+	—	—	—	—
seyen	(+)	+	+	ymb, be, of	to (с XII в.)
tellen		+	+	of	+
speken	(+)		(+)*	ymb, be, of	to, with

* Для глагола *sprecan — speken — to speak* характерны конструкции без объекта. В некоторых случаях этот глагол употребляется с объектом, имеющим характер так называемого «объекта внутреннего содержания».

Условные обозначения:

1. + означает, что данный глагол употребляется в данной конструкции.
2. (+) означает, что хотя данный глагол и употребляется в данной конструкции, все же он выступает в ней реже, чем в конструкции с отметкой +.
3. — означает, что данный глагол не употребляется в данной конструкции.

период

Язык прозы (IX—XI вв.)

Прямая речь	Косвенная речь	Объект	Тема	Имя слушающего
+	+		be	[to] dat.
	+	+	ymb, be	dat.
		(+)*	ymb, be	to dat. wiþ (mid) dat.

Новоанглийский период XVII—XX вв.

	Прямая речь	Косвенная речь	Объект	Тема	Имя слушающего
to say	+	+	+	of	to
to tell		+	(+)	of	+
to speak	(+)		—	of	to, with

4. Незаполненное место означает отсутствие данных.
5. Заштрихованные участки означают, что в данный период глагол находится вне рассматриваемого круга значений. В частности, он может вообще выйти из общего употребления.
6. Конкретное указание союза, предлога, падежа означает употребление глагола в данной конструкции с данным показателем.
7. Формант, находящийся в прямых скобках [ ], может быть пропущен.

употребляется с предлогом *to*, когда речь говорящего направлена на слушающего (в разговоре активно участвует один собеседник), и с предлогом *wiþ*, когда в процессе разговора участвуют оба собеседника.

Глагол *talian* в древнеанглийский период не является глаголом речи. Он имеет значения: 1) «считать», «перечислять» и 2) «считать чем-либо, кем-либо», «полагать» (ср. семантическое отношение в русском языке: «считать до трех» и «считать неправым»). В «Беовульфе» находим второе значение — «считать, полагать», напр.: Swā hyne Gēata bearn gōdne ne tealdon [B. 2184]. 'Так, чтобы дети геатов достойным его не считали'. Первое значение «считать» встретилось в нашем материале в пассивной конструкции с причастием *ʒetealdˊ*. On þam ʒeare sind ʒetealde twelf monðas [AE. T. 30]. 'В году числится 12 месяцев'. Basilissa bið ʒeteald to þære mædena ʒetæle þe marian folʒiað þæs hælendes meder [AE. LS. 94]. 'Василиса пусть будет причислена к числу тех дев, что следуют Марии, матери спасителя'. Примеры. *Cweþan.* Þus cwæd: «Þu þē þyrstende wære monnes blodes XXX wintra, dryncnu þine fylle» [Or. 76]. 'Так сказал: «Ты, который жаждал людской крови тридцать зим, пей теперь ее вволю»'. Sume men cweðað þæt steorran feallaðof heofenum [AE. 1. 66]. 'Некоторые люди говорят, что звезды падают с небес'.

Secʒan. Fela spella him sæðon þa Beormas ʒe of hiera āʒnum lande ʒe of þæm landum þe *ymb* hīe ūtam wæron [Or. 4]. 'Много историй рассказали ему пермяки, как о своей собственной земле, так и о тех землях, что были вокруг них'. Þe mon sæzð þæt on an scip mæʒe an þusend manna [Or. 46]. 'Этот человек говорит, что на один корабль может [сесть] тысяча человек': We wyllað nu ful secʒan þe ðam synfullan wife [AE. LS. 88]. 'Мы хотим теперь все рассказать о грешной женщине'.

Sprecan. He sprecende wæs to his ʒeferum [Or. 84]. 'Он говорил своему спутнику'. Sprecan wiþ þa þe secʒað pæt... [Or. 62]. 'Поговорить с теми, которые говорят, что...'. Him mid spræcon and ða mædena wregdon [AE. LS. 46]. 'Переговорили с ними и ту девственницу обвинили'.

II. Среднеанглийский период

Приблизительно с XIII в. *quothen* употребляется только с прямой речью. К XIV в. остается только форма претерита 3 л. ед. ч. *quod*. К XVII в. *quod* становится малоупотребительным. Существуют среднеанглийские памятники, где *quethen* отсутствует (например из более ранних «Фрагменты речей души к телу» XII в.). Относительно глагола *seyen* следует заметить, что предлоги *ymb* и *bi*, употребляющиеся при теме, к XIII в. сменяются *of* [5].

С XII в. предлог *to* при имени слушающего распространяется на глагол *seyen*. У глагола *tellen* значение «рассказывать», встречающееся первоначально наряду со значением «считать, полагать» и «перечислять, называть», получает все большее распространение [6].

Примеры.

seyen. Þa ʒet seiþ þeo soule soriliche to hire licame: [FRS. 5]. 'И тогда говорит с заботою душа телу'. Goddes ennʒell seʒʒde himm to

[5] Некоторые исследователи связывают распространение *of* на глаголы речи с влиянием старофранцузского (калькирование) (см.: E. Einenkel. Geschichte der Englischen Sprache, II. Historische Syntax. Strassburg. 1916). Однако Оксфордский словарь отмечает случаи употребления *of* с глаголами речи в VIII—IX вв. Ср. также употребление в Беовульфе *fram* в той же функции.

[6] Отдельные случаи употребления глагола *tellen* как глагола речи отмечены в др.-англ. период. Так, Оксфордский словарь указывает на перевод Боэция королем Альфредом (881 г.).

[Orm. 1, 2]. 'Посланец божий сказал ему'. ðo seiden ðis angeles to both wið sped [SGE. 31] 'Тогда ангелы быстро сказали Лоту'.

Tellen. Oc on hem, ðe flogen a — wei Told it abram ðat ilce deai [SGE. 25]. 'А один из тех, что бежали, Поведал это Аврааму в тот же день'. He had wretyn as much as sche wold tellyn hym [MK. 4]. 'Он записал все то, что она пожелала рассказать ему'.

III. Новоанглийский период

К XVII в. между глаголами речи складываются отношения, которые в основном сохраняются до настоящего времени. Это дает возможность не рассматривать материал последующих эпох.

Глагол *quoth* становится малоупотребительным.

Для всей группы в целом характерно единство синтаксического оформления; исключение представляет глагол *to tell*, выступающий с лицом слушающего без *to*. Примеров к этому периоду мы не даем.

Данная выше таблица рабочего материала дает сведения о двух рядах фактов.

1. С одной стороны, следует обратить внимание на тот факт, что в различные периоды истории языка мы имеем дело с одними и теми же наборами синтаксических признаков и одними и теми же значениями, соответствующими этим признакам.

Можно выделить следующие синтаксические признаки и значения (каждую такую группу синтаксических признаков и значений мы будем называть местом).

1) Место для глаголов, стоящих преимущественно при прямой речи, требующих предлога *to* при лице слушающего. Значение глагола: «сказать».

2) Место для глаголов, преимущественно не стоящих при прямой речи, не требующих предлога *to* при лице слушающего. Глаголы на этом месте часто выступают при косвенной речи. Значение глагола: «говорить».

3) Место для глаголов, имеющих те же синтаксические признаки, что и глаголы на месте 2. Отличие от глаголов места 2 лежит в характере объекта [7]. Значение глагола: «рассказывать».

4) Место для глаголов, редко выступающих при прямой и косвенной речи (в отличие от мест 1, 2, 3). Значение: «говорить» (процесс действия).

Сказанное можно резюмировать, воспользовавшись записью из четырех знаков (1 — обозначает наличие признака, 0 — отсутствие признака).

Первый знак (1 или 0) относится к употреблению при глаголе прямой речи, второй знак — косвенной речи, третий знак — объектах с определенным кругом значений, четвертый знак — имени слушающего.

Тогда каждое из мест будет характеризоваться следующими наборами 1 и 0:

1 место	1001	«сказать»
2 место	0100	«говорить»
3 место	0110	«рассказывать»
4 место	0000	«говорить» (процесс)

С другой стороны, обращают на себя внимание явления, связанные с историческим развитием глаголов речи.

На протяжении трех периодов места 1, 2, 3 занимались различными глаголами: 1 место — *maþelian, quethen, to say.* 2 место — *cweþan, seyen, to tell.* 3 место — *secȝan, tellen.* 4 место на протяжении трех периодов было занято одним и тем же глаголом *sprekan, speken, to speak.*

[7] «Рассказывать сказку, историю и т. д.», но «говорить правду, жалкие слова и т. д.» То есть различие между 2 и 3 местами есть различие семантического порядка.

Глаголы, стоявшие на местах 1, 2, 3, проходили единообразный путь развития: к определенному времени глаголы на 1 месте выпадали из языка (*mapelian*, *quethen*), глаголы на 2 месте переходили на 1 место, взамен вышедших из употребления глаголов (*quethen*, *to say*), глаголы на 3 месте переходили соответственно на 2 место (*seyen*, *to tell*).

Можно добавить, что путь развития глагола *talian* 'считать, рассказывать', по-видимому, повторился глаголом романского происхождения *to account* 'рассказывать' (ср. *to count*, *count*). Последний случай не учитывается, однако, в нашей таблице материала.

Вышеизложенные выводы из материала, по-видимому, следует интерпретировать следующим образом.

В языке в области лексики в каждый период существуют определенные семантические модели. Семантические модели такого рода тесно связаны с синтаксическими конструкциями. Семантические модели (и соответствующие им конструкции) могут быть устойчивыми и существовать на протяжении ряда эпох.

Регулярный сдвиг глагольных значений (перемещение глаголов одного «места» на другое) связан, по-видимому, в нашем случае с выпадением глаголов, занимающих место 1. При исчезновении глагола на месте 1 нарушается устойчивая семантическая модель, и в языке проявляется тенденция к восстановлению прежней модели, путем заполнения опустевшего места.

С другой стороны, существуют модели исторического изменения значений. Так, при выпадении глагола с места 1, это место занимается не любым другим глаголом, хотя бы и близким по значению, но непременно глаголом с места 2; освободившееся при этом место 2 занимается глаголом с места 3. Например, глагол sprecan—to speak оказался не принадлежащим к системе мест 1—2—3 и сохранил свое основное значение на протяжении тринадцати столетий.

Перед выпадением из языка глагол речи связан именно с местом 1, определяющим признаком которого является синтаксическая конструкция с прямой речью. По-видимому, прямая и косвенная речь являются синонимичными конструкциями. Прямая речь выражает то же содержание, что и косвенная (т. е. чужую речь), в иной синтаксической форме — в форме бессоюзной связи. При союзной связи — в косвенной речи — для указания на чужую речь существует служебное слово — союз (иногда косвенное наклонение, временной сдвиг и т. д.); при бессоюзии никакого особого указателя начала чужой речи, кроме паузы, не существует. Можно предположить, что употребление одного из глаголов только при прямой речи связано с потребностью своего рода сигнала, указывающего на начало чужой речи. Чтобы закрепиться в позиции при прямой речи, глагол речи проходит длительный путь абстрагирования значения. Как и в случае грамматикализации, глагол имеет тенденцию стать служебным словом, формантом (скажем, типа др.-инд. *iti*). Он может утрачивать морфологические показатели подобно ср.-английскому *quod*, латинскому *ait* или др.-инд. *aha*, *ahur* (pf). Однако глагол при прямой речи формантом не становится; он выходит из общего употребления, иногда сохраняясь за пределами нейтрального стиля. Ср., например, англ. *quoth*, русск. 'молвить, промолвить'.

Взаимоотношение синхронной и исторической моделей можно изобразить в следующей схеме [8]:

---

[8] Глагол *to say* в указанный период еще не перешел окончательно с места 0100 на место 1001.

Место 0110 в указанный период занимает глагол *to account* (романское заимствование), претерпевший тот же семантический сдвиг, что и *talian* —перечислять, считать, рассказывать.

	VIII век	XIII-XIV век	XVII-XX век
СКАЗАТЬ 1001	MAÞELIAN	QUETHEN	TO SAY
ГОВОРИТЬ 0100	CWEÞAN	SEYEN	TO TELL
РАССКАЗЫВАТЬ 0110	SECȝAN	TELLEN	
ГОВОРИТЬ (процесс речи) 0000	SPRECAN	SPEKEN	TO SPEAK

№ 1

СПИСОК ТЕКСТОВ И СОКРАЩЕНИЙ

Древнеанглийский язык

а) Язык эпоса

1. Beowulf, ed. by Claeber, Boston, 1941 (B).

б) Язык прозы

2. King Alfred's Orosius. EETS, № 79, P. 1. 1883 (Or).
3. A E l f r i c. De Temporibus Anni. EETS, № 213, 1942 (AE).
4. AElfric's Lives of Saints. EETS. № 76, P. 1. 1881. (AE. LS).
5. B e d e. The old Endlish version of Ecclesiastical history. EETS, № 95, P. 1. 1890.
6. H o m i l i e s. Blickling. EETS, P. 1. 1874.

Среднеанглийский язык

7. The Ormulum, vol. I, ed. by R. M. White. Oxford 1952 (Orm).
8. Die Fragmenten der Reden der Seele an den Leichname, hsg, von Buchholz. Berlin—Leipzig. 1890 (FRS).
9. The story of Genesis and Exodus. EETS. № 7. P. 1. 1873 (SGE).
10. Das Lied vom King Horn. Strassburg 1881.
11. G o w e r John. The English works. EETS, № 81. Extra series. vol. I, 1900.
12. K e m p e Margere. EETS. 1940 (MK).
13. O s b e r n Bokenham. Legendys of hooly wummen, Lnd, Oxf. 1938.
14. The English Conquest of Ireland, EETS, № 107, 1896.
15. M e r l i n. EETS, № 21, P. 2. Lnd. 1877.
16. L e n g l a n d W. The Vision of William concerning Piers the Plowman. Oxford. 1886.
17. C h a u c e r G. Cannterbury tales. The complete works, vol. IV. Oxf. 1924—35.
18. W y n n e r e and W a s t o u r e. Lnd. 1930.
19. W y c l i f f e John. The English works of Wyclif. Lnd 1902.

Новоанглийский язык

20. S h a k e s p e a r e W. The Tempest. Lnd. 1899.
21. B e h n Aphra. Love-letters between a nobleman and his sister. London, 1735.
22. B u n y a n John. The Pilgrim's progress. Leipzig. 1855.
23. H e a d R. and K i r k m a n. The English Rogue. Lnd. 1665.
24. M i l t o n John. Paradise Lost. N. Y. 1935.

ON THE PATTERN OF VERBA DICENDI IN ENGLISH

Summary

A group of verba dicendi in English (the verbs *to say*, *to tell*, *to account*) reveals a relatively stable pattern that has emerged in the process of language development. The pattern is connected with some syntactical constructions, e. g. direct and indirect speech, various prepositional constructions, etc. This synchronic pattern organises diachronic patterns of meaning change. The regular disappearance of the verb used with the direct speech sets the whole system in motion. Regular shifts of meaning take place.

V. ГРАФЕМИКА И СТРУКТУРНАЯ ТЕОРИЯ ОРФОГРАФИИ

Б. В. Сухотин

ЭКСПЕРИМЕНТАЛЬНОЕ ВЫДЕЛЕНИЕ КЛАССОВ БУКВ С ПОМОЩЬЮ ЭЛЕКТРОННОЙ ВЫЧИСЛИТЕЛЬНОЙ МАШИНЫ [1]

Имеется ряд проблем, решение которых весьма важно для общей и структурной лингвистики. К ним относится, например, вопрос о языковых единицах и о классах языковых единиц. В настоящей работе рассматривается выделение языковых классов применительно к области, традиционно именуемой фонологией, если правомерно включить в фонологию изучение букв.

Говоря о классах звуков, имеют в виду прежде всего гласные и согласные. Можно ли говорить о подобных классах применительно к буквам? Какие свойства звуков остаются неизменными при замене звуков буквами? Мы не видим других общих свойств, кроме более или менее сходной сочетаемости звуков и букв. Поэтому только на этой основе нам представляется целесообразным строить деление букв на классы.

Так, представляется почти очевидным, что за гласными обычно следуют согласные, за согласными гласные, а если кое в каких языках, кое в каких орфографиях это не так, то из этого предположения хорошо исходить, чтобы узнать, какие дистрибутивные закономерности там налицо.

Очевидно также, что эта закономерность — неточная, т. е. бывает, что гласные следуют за гласными и согласные — за согласными. Поэтому трудность классификации возрастает. Кроме того, затрудняет классификацию разная частота букв. Например, сочетаний с согласными для частой русской буквы *с* может оказаться больше, чем сочетаний с согласными для редкой буквы *э*. Поэтому необходим такой способ, в котором все эти усложнения были бы учтены.

Основным нашим инструментом будет таблица частот двубуквенных сочетаний. Опишем эту таблицу.

Таблица частот двубуквенных сочетаний — квадратная; строчек и столбцов в ней столько, сколько букв в данном алфавите; каждый столбец озаглавлен той же буквой, что и строка с соответствующим номером; номера можно считать от верхнего левого угла таблицы. Правила заполнения этой таблицы такие:

1. На пересечении строчки x и столбца y отмечается число сочетаний, имеющих в качестве первого элемента x, в качестве второго — y. Вообще говоря, ввиду того, что гласные, например, очевидно, не только следуют за согласными, но и предшествуют им, можно не делать разницы между следованием и предшествованием и считать, что $xy=yx$. Тогда на пересе-

[1] Автор выражает глубокую благодарность С. К. Шаумяну, К. И. Бабицкому (Институт русского языка АН СССР), И. И. Новикову (Одесский университет им. И. И. Мечникова), Т. М. Великановой, Н. М. Нагорному и А. В. Кузьминой (ВЦ АН СССР) за помощь и поддержку, которые трудно переоценить.

чении строчки x и столбца y надо отмечать сумму сочетаний с x в качестве первого элемента, а y — в качестве второго и сочетаний с y в качестве первого элемента, а x — в качестве второго. Такая таблица будет симметричной относительно главной диагонали.

2. Считается, что за последней буквой некоторого слова следует первая буква следующего слова и за последней буквой текста следует первая буква текста, т. е. пробел не учитывается. Тем самым достигается большая общность постановки задачи, так как есть орфографии, не имеющие пробела, а кроме того, пробел нельзя разумным образом интерпретировать как гласный или согласный.

Как отразится на такой таблице предполагаемая закономерность — именно, что за гласными часто следуют согласные и редко — гласные, а за согласными часто следуют гласные и редко — согласные? Представим себе, что строчки и столбцы, озаглавленные гласными, отодвинуты в левый верхний угол таблицы, тогда таблица примет следующий вид:

	Гласные	Согласные
Гласные	1	2
Согласные	4	3

Таблица разобьется на четыре части линиями, отделяющими гласные от согласных. В квадрате 1 будут отмечены сочетания гласных с гласными, в квадрате 3 — согласных с согласными, в прямоугольниках 2 и 4 — сочетания гласных с согласными. Если наше предположение верно, то сумма отметок в квадрате 1 будет невелика, в квадрате 3 — тоже, а в прямоугольниках 2 и 4 сумма отметок должна быть большой. Если мы переставим местами строчки и столбцы, озаглавленные только гласными или только согласными, то все эти суммы, конечно, не изменятся, так как слагаемые останутся теми же, но если мы к согласным присоединим какую-нибудь гласную, или наоборот, или поменяем местами гласную и согласную, то суммы изменятся. Легко заметить, что при любом таком изменении сумма отметок в квадратах 1 и 3 не должна уменьшиться, а сумма отметок в прямоугольниках 2 и 4 не должна увеличиться. Пусть, например, мы включили какую-нибудь согласную в гласные. Тогда сочетания этой согласной с гласными попадут в квадрат 1, а сочетания этой согласной с согласными выйдут из квадрата 3. Но сочетаний согласной с гласными, по нашему предположению, больше, чем сочетаний согласной с согласными, поэтому общая сумма отметок в квадратах 1 и 3 должна увеличиться. Поэтому, если гласные правильно отделены от согласных, то сумма отметок в квадратах 1 и 3 должна быть не только небольшой, но и наименьшей. Соответственно, в прямоугольниках 2 и 4 сумма отметок должна быть наибольшей.

Поэтому ясно, что достаточно получить такое расположение строк и столбцов и такое деление таблицы двумя линиями, чтобы в квадратах 1 и 3 сумма отметок была минимальной, как сразу же мы получим требуемое деление букв нашего алфавита на два класса.

К описанию такого способа мы сейчас и переходим. Основная идея этого способа заключается в следующем: предположим, что мы уже отодвинули в левый верхний угол какие-нибудь строчки и столбцы, принадлежащие к одному из классов. Мы провели границу, отделяющую уже перенесенные строчки и столбцы от неперенесенных. Нам нужно узнать, надо ли переносить в этот угол, за границу, какую-нибудь новую строчку и столбец, и если надо, то какую.

Вычислим для каждой строчки, стоящей н и ж е границы, разность суммы отметок с п р а в а от вертикальной черты и суммы отметок, стоящих с л е в а от вертикальной черты. Если эта разность отрицательна, то строчку и столбец, озаглавленные данной буквой, не следует переносить в левый верхний угол. В самом деле, ведь это значит, что правая часть строчки меньше (по сумме отметок) левой части и перенесение строчки наверх, через границу, увеличит сумму отметок в квадратах 1 и 3. То же *самое справедливо и для соответствующего столбца.*

Теперь выясним, какую из строчек, обладающих положительной разностью правых и левых частей, нужно переносить в первую очередь. Посмотрим сначала, какое влияние оказывает перенос строчки (и столбца) на разности строк, оставшихся ниже границы. Квадрат 1 в результате перенесения увеличивается, а квадрат 3 уменьшается: следовательно, в правой части строчек суммы будут состоять из меньшего на единицу числа слагаемых, а в левой части строчек суммы будут состоять из большего на единицу числа слагаемых.

Следовательно, разности будут стремиться к уменьшению (не по абсолютной величине!).

Поэтому, если при данном положении границы какая-то строчка обладает наибольшей положительной разностью, надо «использовать благоприятный момент» и перенести ее и соответствующий столбец через границу; ведь большей разности уже не будет, а нам нужны максимальные разности между квадратами (тем самым, строчками квадратов) 1 и 3, 4 и 2.

Итак, мы знаем, что если некоторые строчки (соотв. столбцы) уже перенесены, то перенести надо строчку (и столбец) с максимальной положительной разностью; если все разности отрицательны, то переносить ничего больше не надо, процесс перенесения строк и столбцов окончен.

Теперь достаточно принять, что верхняя и левая рамки таблицы суть границы, отделяющие перенесенные строчки и столбцы от еще не перенесенных (число перенесенных строчек и столбцов равно нулю), чтобы завершить — по индукции — построение нашего способа.

Перейдем к рассмотрению некоторых второстепенных деталей. Так, разность правой и левой части данной строчки может оказаться равной нулю. Такую строчку, вообще говоря, можно переносить и не переносить; мы ее, для единообразия, не переносим. Может оказаться, что две или несколько максимальных разностей равны. В этом случае следует перепробовать все варианты, и если они не совпадают, выбрать тот, при котором суммы в квадратах 1 и 3 убывают быстрее. Практически наличие нескольких максимальных разностей — вещь весьма редкая, особенно при достаточно большом тексте, поэтому тут можно пойти на упрощение и просто испытывать один из вариантов.

Следует напомнить, что способ излагался применительно к симметричной таблице. В случае несимметричной таблицы необходимо внести следующее изменение: вместо «левой» суммы надо взять сумму отметок

на данной строчке левее вертикальной границы плюс сумму отметок на соответствующем столбце выше горизонтальной границы; вместо «правой» суммы — сумму отметок на данной строчке правее вертикальной границы плюс сумму отметок на соответствующем столбце ниже горизонтальной границы.

Отметки, стоящие на главной диагонали, всегда входят в квадраты 1 и 3, поэтому их лучше не учитывать — они не влияют на целесообразность переноса строк и столбцов.

После проведения подсчета по этому способу, квадраты 1 и 3 можно рассматривать как новые таблицы и снова делить их на четыре части, получая, таким образом, деление каждого класса на подклассы. Если на подклассы мы делим гласные, то подклассами будут основные и неслоговые элементы дифтонгов; в случае согласных, очевидно, смычные, щелевые, сонорные и т. д.

При сравнении результатов такого деления с принятым делением на гласные и согласные необходимо учитывать разницу между орфографией и произношением.

А сейчас мы еще раз запишем наш способ в виде последовательности правил.

1. В квадратной таблице порядка n строчки и столбцы озаглавливать буквами алфавита в произвольном порядке, однако так, чтобы столбец с номером N был озаглавлен так же, как и строка с тем же номером.

2. Записать в клеточке на пересечении строчки x и столбца y число, означающее, сколько раз x и y встретились в непосредственном соседстве.

3. Зачеркнуть числа, стоящие на главной диагонали.

4. Провести границы, отделяющие перенесенные строчки и столбцы от неперенесенных.

5. Для каждой строки ниже горизонтальной границы подсчитать разность суммы отметок на пересечении данной строки с еще не перенесенными столбцами и суммы отметок на пересечении данной строки с уже перенесенными столбцами.

6. Среди строчек, расположенных ниже границы, выбрать строчку с максимальной положительной разностью и перенести ее через границу. То же проделать с соответствующим столбцом. Если среди проверяемых строк нет положительных, процесс закончить. Если есть, выполнить правило 4.

7. Если есть несколько максимальных разностей, то проверить все варианты и выбрать тот, для которого сумма отметок в квадратах, стоящих на диагонали, убывает быстрее, в противном случае дать все ответы, если они не совпадают.

Этот способ был запрограммирован И. И. Новиковым (Одесса) для машины БЭСМ-2, находящейся в ВЦ АН СССР, со следующими упрощениями: 1. Условие (7) не учитывалось. 2. Числа на главной диагонали не зачеркивались.

Для первоначальных опытов А. В. Кузьминой были закодированы тексты на русском, английском и французском языках, объемом в 10 000 букв каждый. Объем текста выбран на основании рекомендаций в пособиях по криптографии.

Таблицы с вертикальной и горизонтальной границами, отделяющими гласные от согласных, которыми озаглавлены строчки и столбцы, приведены ниже. Они представляют результат подсчетов, проделанных машиной по указанному способу.

Рассмотрим полученные результаты. В русском тексте в один класс попали *о*, *а*, *е*, *и*, *у*, *ъ*, *ь*, *я*, *ы*, *э*, *ю*, в другой — *б*, *в*, *г*, *д*, *ж*, *з*, *й*, *к*, *л*, *м*, *н*, *п*, *р*, *с*, *т*, *ф*, *х*, *ц*, *ч*, *ш*, *щ*. Не следует удивляться тому, что *я* и *ю* попали в гласные — они являются гласными с точки зрения орфографии, так как часто встречаются в окружении согласных. Естественно, что ь — исторически

		Гласные											Согласные																											
		О	А	Е	И	У	Ъ	Ь	Я	Ы	Э	Ю	Б	В	Г	Д	Ж	З	Й	К	Л	М	Н	П	Р	С	Т	Ф	Х	Ц	Ч	Ш	Щ
Гласные	О	28	15	39	32	7	4	3	16	2	3	7	72	167	136	107	24	30	56	143	192	118	181	152	121	133	192	6	44	4	31	25	6
	А	15	6	10	22	10	2	1	30	1	2	13	35	127	43	99	40	90	10	147	173	67	228	46	120	103	110	11	42	9	60	50	4
	Е	39	10	28	57	7	5	15	5	19	4	4	46	83	37	93	44	35	34	43	166	87	202	71	156	129	124	0	16	19	63	22	27
	И	32	22	57	52	13	9	2	36	4	2	2	23	95	25	60	26	37	31	83	148	102	143	42	68	89	95	5	28	18	43	39	21
	У	7	10	7	13	0	3	2	4	1	3	19	20	20	26	42	12	11	0	41	24	26	43	27	42	34	53	0	10	2	16	16	8
	Ъ	4	2	5	9	3	0	0	11	0	3	6	5	21	2	12	0	4	0	20	42	9	40	8	9	70	59	0	0	2	10	7	1
	Ь	3	1	15	2	2	0	0	1	0	0	0	26	27	1	8	0	6	11	4	22	31	43	7	14	7	15	0	12	4	1	5	0
	Я	16	30	5	36	4	11	1	8	0	0	0	8	31	6	15	4	6	1	10	30	10	46	14	11	60	31	1	4	0	1	1	1
	Ы	2	1	19	4	1	0	0	0	0	0	0	14	13	1	11	0	4	10	3	13	15	37	8	8	5	14	0	6	3	0	4	0
	Э	3	2	4	2	3	3	0	0	0	0	0	0	4	0	0	0	0	1	3	2	0	0	1	0	1	23	0	0	0	0	0	0
	Ю	7	13	4	2	19	6	0	0	0	0	2	5	3	0	2	0	3	0	2	9	1	2	4	1	5	13	0	0	0	1	0	14
Согласные	Б	72	35	46	23	20	5	26	8	14	0	5	0	2	0	1	0	1	2	7	12	0	3	0	9	5	2	0	4	0	0	0	2
	В	167	127	83	95	20	21	27	31	13	4	3	2	8	5	24	1	20	5	17	16	16	16	17	21	44	21	0	8	4	3	21	0
	Г	136	43	37	25	26	2	1	6	1	0	0	0	5	0	19	0	5	4	3	16	6	7	1	19	1	2	8	1	0	0	0	0
	Д	107	99	93	60	42	12	8	15	11	0	2	1	24	19	2	13	10	2	4	9	5	39	4	24	7	10	0	0	8	0	0	0
	Ж	24	40	44	26	12	0	0	4	0	0	0	0	1	0	13	0	2	0	12	4	0	17	0	10	1	0	0	0	0	0	0	0
	З	30	90	35	37	11	4	6	6	4	0	3	1	20	5	10	2	0	3	6	10	14	20	3	5	2	6	0	2	1	0	0	0
	Й	56	10	34	31	0	0	11	1	10	1	0	2	5	4	2	0	3	0	11	6	4	14	12	8	23	13	0	3	1	0	3	0
	К	143	147	43	83	41	20	4	10	3	3	2	7	17	3	4	12	6	11	4	19	8	11	10	40	57	15	0	5	0	3	5	0
	Л	192	173	166	148	24	42	22	30	13	2	9	12	16	16	9	4	10	6	19	2	10	13	21	4	43	8	0	9	0	5	12	0
	М	118	67	87	102	26	9	31	10	15	0	1	0	16	6	5	0	14	4	8	10	16	23	10	12	21	5	0	2	1	3	0	0
	Н	181	228	202	143	43	40	43	46	37	0	2	3	16	7	39	17	20	14	11	13	23	80	7	12	32	20	0	9	8	12	2	2
	П	152	46	71	42	27	8	7	14	8	1	4	0	17	1	4	0	3	12	10	21	10	7	0	59	17	8	0	6	0	2	0	1
	Р	121	120	156	68	42	9	14	11	8	0	1	9	21	19	24	10	5	8	40	4	12	12	59	0	16	48	0	20	1	2	4	0
	С	133	103	129	89	34	70	7	60	5	1	5	55	44	1	7	1	2	23	57	43	21	32	17	16	16	160	0	6	0	12	1	0
	Т	192	110	124	95	53	59	15	31	14	23	13	2	21	2	10	0	6	13	15	8	5	20	8	48	160	4	1	3	0	22	0	1
	Ф	6	11	0	5	0	0	0	1	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	Х	44	42	16	28	10	0	12	4	6	0	0	4	8	1	0	0	2	3	5	9	2	9	6	20	6	3	0	0	0	0	2	0
	Ц	4	9	19	18	2	2	4	0	3	0	0	0	4	0	8	0	1	1	0	0	1	8	0	1	0	0	0	0	0	1	0	0
	Ч	31	60	63	43	16	10	1	1	0	0	1	0	3	0	0	0	0	0	3	5	3	12	2	2	12	22	0	0	1	0	1	0
	Ш	25	50	22	39	16	7	5	1	4	0	0	0	21	0	0	0	0	3	5	12	0	2	0	4	1	0	0	2	0	1	0	0
	Щ	6	4	27	21	8	1	0	1	0	0	14	2	0	0	0	0	0	0	0	0	0	2	1	0	0	1	0	0	0	0	0	0

		Гласные							Согласные																	
		e	o	a	i	t	u	y	b	c	d	f	g	h	j	k	l	m	n	p	q	r	s	v	w	x	z
Гласные	e	116	48	112	63	108	8	52	70	65	204	57	56	430	1	35	135	109	154	61	1	312	170	111	106	7	1
	o	48	114	15	20	157	115	17	29	33	71	97	41	91	0	38	119	70	125	30	0	139	80	14	91	0	0
	a	112	15	6	48	138	4	44	24	40	89	27	64	97	0	16	104	87	214	34	1	105	170	30	90	1	1
	i	63	20	48	0	102	9	8	5	26	73	37	39	140	0	23	88	48	203	64	8	119	121	23	53	0	1
	t	108	157	138	102	104	57	32	8	13	60	35	20	352	1	7	43	16	119	19	0	77	139	1	28	0	0
	u	8	115	4	9	57	0	0	32	9	13	7	34	12	3	0	23	14	63	14	12	31	52	0	0	0	0
	y	52	17	44	8	32	0	2	18	5	17	8	5	20	0	2	42	16	13	5	1	33	25	3	19	0	1
Согласные	b	70	29	24	5	8	32	18	4	0	15	1	3	6	1	1	7	6	11	1	0	18	13	0	1	0	0
	c	65	39	40	26	13	9	5	0	0	5	1	1	46	0	11	8	1	16	1	0	14	23	0	1	1	0
	d	204	71	89	73	60	13	17	15	5	22	11	12	49	2	0	63	20	155	5	1	44	39	1	17	0	0
	f	57	97	27	37	35	7	8	1	1	11	16	4	12	0	2	27	2	5	4	0	18	20	0	7	0	0
	g	56	41	64	39	20	34	5	3	1	12	4	6	45	0	1	2	4	106	1	1	31	25	0	9	0	0
	h	430	91	97	140	352	12	20	6	46	49	12	45	16	0	3	4	5	27	8	0	23	62	0	66	0	0
	j	1	0	0	0	1	3	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	k	35	38	16	23	7	0	2	1	11	0	2	1	3	0	0	13	0	26	1	0	7	15	0	1	0	0
	l	135	119	104	88	43	28	42	7	8	63	27	2	4	0	13	138	8	16	33	0	7	30	3	15	0	0
	m	109	70	87	48	16	14	16	6	1	20	2	4	5	0	0	8	2	5	6	0	11	17	1	8	0	0
	n	154	125	214	203	119	63	13	11	16	155	5	106	27	0	26	16	5	20	0	0	17	44	1	32	0	0
	p	61	30	34	64	19	14	5	1	1	5	4	1	8	0	1	33	6	0	26	0	20	23	13	10	5	0
	q	1	0	1	8	0	12	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0
	r	312	139	105	119	77	31	33	18	14	44	18	31	23	0	7	7	11	17	20	1	46	45	2	10	0	0
	s	170	80	170	121	139	52	25	13	23	39	20	25	62	0	15	30	17	44	23	1	45	74	7	41	0	0
	v	111	14	30	23	1	0	3	0	0	1	0	0	0	0	0	3	1	1	13	0	2	7	0	0	0	0
	w	106	91	90	53	28	0	19	1	1	17	7	9	66	0	1	15	8	32	10	1	10	41	0	6	0	0
	x	7	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
	z	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Гласные		e	a	o	i	u	y	k	Согласные b	c	d	f	g	h	j	l	m	n	p	q	r	s	t	v	w	x	z
Гласные	e	96	67	6	114	166	6	0	40	168	283	33	45	46	30	318	219	302	127	27	428	457	283	109	0	14	12
	a	67	16	0	116	90	22	1	27	63	81	34	37	25	8	173	72	166	127	5	194	144	118	89	0	7	1
	o	6	0	0	118	150	3	0	35	73	19	14	2	22	10	43	71	253	60	0	118	88	64	43	0	1	0
	i	114	116	118	4	68	1	4	13	45	43	17	12	8	5	137	55	100	17	5	85	147	140	39	0	10	1
	u	166	90	150	68	4	0	0	9	25	57	3	10	3	7	70	10	92	17	94	111	146	83	20	0	33	0
	y	6	22	3	1	0	0	0	0	4	1	0	0	0	0	11	11	3	7	0	0	7	2	0	0	0	0
	k	0	1	0	4	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	3	0
Согласные	b	40	27	35	13	9	0	0	0	0	0	0	0	0	0	21	4	2	0	0	13	11	5	0	0	2	0
	c	168	63	73	45	25	4	1	0	16	2	0	0	56	0	15	1	55	3	0	27	37	36	0	0	3	2
	d	283	81	19	43	57	1	1	0	2	12	1	1	0	1	1	1	83	5	2	64	69	37	2	0	5	0
	f	33	34	14	17	3	0	0	0	0	1	10	0	0	0	11	0	11	0	0	7	5	3	0	0	1	0
	g	45	37	2	12	10	0	0	0	0	1	0	0	0	0	2	0	22	0	1	18	3	3	0	0	0	0
	h	46	25	22	8	3	0	0	0	56	0	0	0	0	0	8	0	4	2	0	1	8	1	0	0	2	0
	j	30	8	10	5	7	0	0	0	0	1	0	0	0	0	1	0	2	0	0	1	5	4	0	0	0	2
	l	318	173	43	137	70	11	0	21	15	1	11	2	8	1	74	3	24	28	7	36	88	44	3	0	2	2
	m	219	72	71	55	10	11	0	4	1	1	0	0	0	0	3	62	5	19	0	12	9	11	0	0	1	0
	n	302	166	253	100	92	3	0	2	55	83	11	22	4	2	24	5	60	14	14	16	100	170	12	0	0	0
	p	127	127	60	17	17	7	0	0	3	5	0	0	2	0	28	19	14	22	0	58	49	28	0	0	7	2
	q	27	5	0	5	94	0	0	0	0	2	0	1	0	0	7	0	14	0	0	8	20	6	0	0	1	0
	r	428	194	118	85	111	0	1	13	27	64	7	18	1	1	36	12	16	58	8	96	58	160	10	0	0	0
	s	457	144	88	147	146	7	1	11	37	69	5	3	8	5	88	9	100	49	20	58	152	112	11	0	1	0
	t	283	118	64	140	83	2	0	5	36	37	3	3	1	4	44	11	170	28	6	160	112	54	10	0	1	0
	v	109	89	43	39	20	0	0	0	0	2	0	0	0	0	3	0	12	0	0	10	11	10	0	0	2	2
	w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	x	14	7	1	10	33	0	3	2	3	5	1	0	2	0	2	1	0	7	1	0	1	1	2	0	0	0
	z	12	1	0	1	0	0	0	0	2	0	0	0	0	2	2	0	0	2	0	0	0	0	2	0	0	0

гласный — по той же причине попал в гласные. Знак ъ попал в гласные в результате ошибки при кодировании текста. Во французском тексте в один класс попали *e*, *a*, *o*, *i*, *u*, *y*, *k*, в другой *b*, *c d*, *f*, *g*, *h*, *j*, *l*, *m*, *n*, *p*, *q*, *r*, *s*, *t*, *v*, *w*, *x*, *z*. Буква *k*, попавшая в гласные, встретилась всего 6 раз, видимо, в особых условиях, так как в исконно французских словах эта буква не встречается. Буква *w*, не встретившаяся ни разу, попала в согласные, как обладающая нулевой разностью.

При анализе английского текста в один класс попали *e*, *o*, *a*, *i*, *t*, *u*, *y*, в другой — *b*, *c*, *d*, *f*, *g*, *h*, *j*, *k*, *l*, *m*, *n*, *p*, *q r*, *s*, *v*, *w*, *x*, *z*. Здесь интересно то, что *t* попала в гласные. Причина этого, конечно, в том, что частое сочетание *th* нередко следует после согласного, ср. *of the*. . . В настоящее время программируется исправляющий алгоритм для устранения ошибок подобного рода. Очевидно, при дистрибутивном выделении классов можно достигнуть значительной точности.

При оценке полученных результатов нужно отметить следующее: наш способ позволяет разбить алфавит на два класса, не указывая, какой из них отражает гласные, какой согласные, и т. д.

Существует мнение, что такие дистрибутивные признаки, которые бы позволили квалифицировать данный класс именно как гласные (согласные) отсутствуют. Нам кажется, что это мнение ошибочно. В качестве эвристического указания можно, например, воспринять тот факт, что класс, включающий максимально частый знак, возможно, соответствует гласным; если это не так или не совсем так, то интересно, до какой степени это не так.

В заключение нужно отметить, что способы подобного рода могут быть весьма многочисленными и найти применение в ряде других лингвистических задач. Так, например, если почти такой же способ применить не для букв, а для иероглифов, то можно ожидать, что мы получим разделение иероглифов на классы существительных и глаголов или чего-нибудь в этом роде.

EXPERIMENT IN CLASSIFYING LETTERS BY COMPUTOR

Summary

In this paper an attempt is made to classify letters according to their distribution. It is supposed that only distribution offers a basis for a classification of letters which would not be senseless in terms of sounds. The following algorithm was employed for this purpose:

1. Form a square matrix with columns and lines corresponding to the letters of a given alphabet. The line corresponding to a certain letter must be of the same number as the column corresponding to the same letter.

2. At the intersection of line *x* and column *y* write the frequency of the combination *xy* or *yx* in the given text.

3. Cross out all frequencies on the main diagonal.

4. Draw border-lines separating transferred rows from not transferred ones.

5. For each line subtract the sum of the frequencies to the left of the vertical border-line from the sum of the frequencies to the right of the vertical border-line.

6. Among lines below the horizontal border-line take the line with the maximal positive difference and transfer it over the border-line. Do the same with the column of the same number. If there are no more positive differences among those below the horizontal border-line, finish the process. If there are such, do N 4.

This algorithm was programmed for БЭСМ-2 by I. I. Novikov (Odessa). Three texts (Russian, English, French) were chosen for the analysis. Each text comprised 10 000 letters. The results are as follows.

1. Russian. Vowels (or class I): *о, а, е, и, ъ, ь, ы, э, ю, я*; Consonants (or class II): *б, в, г, д, ж, з, й, к, л, м, н, п, р с, т, ф, х, ц, ч, ш, щ*.

2. English. Vowels (or class I): *e, o, a, i, t, u, y*. Consonants (or class II): *b c, d, f, g, h, j, k, l, m, n, p, q, r, s, v, w, x, z*.

3. French. Vowels (or class I): *e, a, o, i, u, y, k*. Consonants (or class II): *b, c, d, f, g, h, j, l, m, n, p, q, r, s, t, v, w, x, z*.

For correction of mistakes which are not accidental (such as with the English *t*) a correction algorithm has been worked out and is being programmed.

We hope that considerable precision may be achieved in such methods.

В. Г. Гак

ОРФОГРАФИЯ В СВЕТЕ СТРУКТУРНОГО АНАЛИЗА

(на материале французского языка)

> Все, в чем выражено данное состояние языка, надо уметь свести к теории синтагм и к теории ассоциаций.
>
> *Ф. де Соссюр*

Графическая система языка резко отличается от его фонологической, грамматической или лексической систем. Она является в значительной степени внешней по отношению к структуре языка, гораздо более подвержена различным условным изменениям (в истории известны случаи, когда на протяжении жизни одного поколения система письменности у некоторых языков менялась дважды). Однако графика и орфография, подобно иным подсистемам языка, социально обусловлены и покоятся на тех же общих семиотических законах, что и система фонем или морфем языка.

Структурный анализ системы письменности представляет, следовательно, двоякий интерес [1]. Во-первых, этот анализ дает возможность научно выявить закономерности, определяющие функционирование системы графики и орфографии данного языка. Во-вторых, этот анализ может быть интересен и с обратной точки зрения — он может способствовать углублению самой семиотики, выявлению общих закономерностей функционирования знаковых систем на всех уровнях языка. Для семиотики изучение письменности может иметь большое значение. Письменность представляет собой систему знаков в чистом виде. Здесь связь между означающим и означаемым наиболее непосредственна и не осложнена многообразными семантическими наслоениями, как это наблюдается в грамматике и в лексике (не случайно Ф. де Соссюр во многих случаях для иллюстрации положений своей теории знака прибегает к примерам из орфографии). Различные явления и закономерности семиотики, рельефно проявляющиеся в системе письменности, имеют свои более завуалированные аналогии в морфологии, синтаксисе, словообразовании, фразеологии и других разделах языка.

---

[1] Попытки применения методов структурного анализа в изучении графики и орфографии см., например, в кн.: Г. Глиссон. Введение в дескриптивную лингвистику. М., 1959, главы XXI, XXII; Опыт определения системы графем на основе дистрибутивного анализа (для датского языка) — в кн.: H. Spang-Hanssen. Probability and structural classification in Language description. Copenhagen, 1959, главы IV, V. Интерес к структурному анализу графических систем разных языков стимулируется в последнее время также потребностями машинного перевода.

Исследуя французскую орфографию в свете структурного анализа, мы ставим себе задачей в этой статье рассмотреть следующие вопросы:

1. Особенности графем языка.
2. Методы структурной лингвистики и их применение к изучению орфографии.
3. Определение понятия орфографического правила с точки зрения структурного анализа.

ОСОБЕННОСТИ ГРАФЕМ ЯЗЫКА

Графическая система состоит из определенного количества графических единиц — графем. Как и всякая знаковая единица графема представляет собою единство означающего и означаемого [2].

С материальной точки зрения означающим может быть определенного начертания буква, буквосочетание, орфографический знак (апостроф, дефис, надстрочные или подстрочные знаки и т. п.). Означаемым содержанием знака обычно выступает фонема, но иногда содержанием знака является некоторая функция в системе письменности, принятая в данном языке, причем знак лишь косвенным образом соотносится со звучанием. Таковы ь или ъ в русской графике, которые сами по себе никакого звука не обозначают, но дифференцируют значение других многозначных знаков.

Графемы, как и другие знаки языковой системы, могут иметь разные в а р и а н т ы: позиционные, функциональные, индивидуальные и национальные.

П о з и ц и о н н ы е варианты графем суть те, которые зависят от места данного знака в слове. Так, в арабском алфавите буквы, обозначающие многие согласные звуки, имеют различные начертания в зависимости от их позиции по отношению к другим буквам. В греческом, готическом и старинном французском шрифте буква *s* писалась по-разному в середине и в конце слова.

Позиционные варианты находятся в отношении дополнительной дистрибуции и образуют аллографы.

Ф у н к ц и о н а л ь н ы е варианты наиболее разнообразны. Здесь можно указать на прописные и заглавные буквы, на различные виды шрифтов (печатный, рукописный, прямой, курсив, афишный и т. д.), подчас значительно видоизменяющие внешний вид знаков.

И н д и в и д у а л ь н ы е варианты также весьма многообразны. Одни пишут букву d *d*, другие *∂*, по-разному изображается разными лицами z (*z* и *ẑ*) и некоторые другие буквы.

И, наконец, в том случае, когда одним алфавитом пользуются разные языки, можно говорить о н а ц и о н а л ь н ы х вариантах одних и тех же знаков. Так, по-разному пишут заглавные *T*, *I*, *G* и другие буквы англичане, французы, немцы.

Данное начертание определяется как вариант графемы лишь на основании связи с означаемым. В русском письме знаки *∂* и *д* имеют значительно большие материальные отличия, нежели знаки *и* и *ц*. И все же первые два — варианты одного знака д, а два других — разные знаки (и и ц) именно потому, что вторые два регулярно обозначают различные звуки, а первые — регулярно обозначают один звук. Объединение вариантов

---

[2] В некоторых работах (см., например: С п а н г - Х а н с с е н. Указ. соч.) графема выделяется на основании дистрибутивного анализа письменной речи, независимо от ее референта в устном языке. Для орфографии графемы имеют значение прежде всего потому, что они изображают элемент звукового языка. Выделение графики должно, следовательно, основываться не только на изучении внутренней структуры системы графики, но и на анализе соответствия между буквой и фонемой, между письменным и устным языком (см.: Г л и с с о н. Указ. соч., стр. 399).

в одну графему определяется также не близостью их материальной формы, но тождеством означаемого. Так, если мы возьмем три знака: *M m T*, то обнаружим, что в графической системе французского языка среднее начертание принадлежит к той же графеме, что и *M*, в то время как в русской этот знак есть вариант той же буквы, что и *T*.



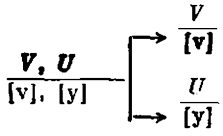
Во французской графике *g* и *д* суть разные знаки, в русской — это вариант одной графемы.

Возможность изменения формы знака определяется системой. Такое изменение нежелательно, если приводит к смешению данного знака с другими.

В истории орфографии известны случаи, когда варианты одного знака превратились в два разных знака и наоборот. В основе этой эволюции лежит изменение соотношения между означаемым и означающим. Так в старофранцузской письменности написания *U* и *V* были вариантами одной и той же буквы, которая имела два основных значения: гласное [y] и согласное [v]. В дальнейшем эти написания дифференцировались: первое закрепилось за гласным звуком, второе стало обозначать только согласный звук.

Так образовались два отдельные знака.

Эволюция шла по схеме:



Аналогичным образом варианты графемы *i* и *j* превратились в две разных графемы.

С другой стороны, может наблюдаться тенденция к превращению отдельных графем в аллографы и, следовательно, в варианты одной графемы. Так, в романских языках буквы *c* и *g* находятся в отношении дополнительной дистрибуции с другими знаками.

Язык	Фонема	Перед e, i [y]	В остальных случаях	Фонема	Перед e, i [y]	В остальных случаях
Итальянский	[k] [tʃ]	ch c	c ci	[g] [dʒ]	gh g	g gi
Испанский	[k] [θ]	qu c	c z	[g] [x]	gu g	g j
Французский	[k] [s]	qu c	c ç	[g] [ʒ]	gu g	g ge

Однако подлинными аллографами эти знаки считать нельзя. Этому препятствуют три момента, нарушающие дополнительную дистрибуцию:

а) многозначность знаков *c* и *g* во всех языках;

б) наличие в графической системе языка других знаков, изображающих ту же фонему, независимо от указанной дистрибуции; например,

буква *k*, обозначающая [k] во всех трех языках, буквы *s* и *j*, обозначающие фонемы [s] и [ʒ] во французском языке;

в) непоследовательность некоторых из указанных дистрибуций; так в некоторых испанских словах *z* пишется и перед *i* (*zigzag*, *zipizape*), франц. *qu* пишется часто перед *a* (*quand*) и т. п. Если бы не эти факты, то, например, в испанском языке буквы *z* и *c*, несмотря на их внешнее различие и различное место в алфавите, должны были бы рассматриваться не как разные графемы, а как аллографы одной и той же графемы.

В качестве аллографов во французской орфографии выступают *e* [ɛ] и *è* [ɛ]. Первый знак встречается только в закрытом слоге (*il trompette*), а второй — в открытом (*il trompète*). В отношении дополнительной дистрибуции находятся знаки: *s* [s] (встречается во всех позициях, кроме позиции между двумя гласными) и *ss* (встречается только между двумя гласными). Ср. *sieste* и *basse* [3]. Двойная и одинарная согласная выступают как аллографы в некоторых позициях. Так на конце слова, после гласной буквы с accent и после согласной буквы может быть только одна согласная, двойная согласная может быть только в позиции между двумя гласными, из которых первая не несет accent. Ср. *rôle* и *colle*, *fer* и *ferreux*, *panse* и *passe* [4].

Идеальная система знаков должна была бы включать в себя столько графем-букв, сколько есть фонем в языке, причем между звуками и буквами должно быть взаимно-однозначное соответствие. Графема, характеризующаяся таким соотношением, может быть названа абсолютной. Такие абсолютные знаки в современных графических системах — большая редкость. Так, в русской орфографии абсолютным знаком является *у* (*у* всегда обозначает звук [у] и [у] всегда изображается буквой *у*). Во французской графике абсолютных знаков нет.

Итак, каждый а б с о л ю т н ы й знак (графема) характеризуется о т д е л ь н о с т ь ю и о д н о з н а ч н о с т ь ю. Редкость абсолютного знака проистекает вследствие нарушения одного из этих качеств. А это нарушение в свою очередь объясняется наличием в языке с и н т а г м а т и ч е с к и х и п а р а д и г м а т и ч е с к и х (а с с о ц и а т и в н ы х) о т н о ш е н и й.

Нарушение принципа о т д е л ь н о с т и знака связано с действием с и н т а г м а т и ч е с к и х связей между различными означаемыми и означающими. Здесь могут быть два случая.

1. Одно означающее (буква) соответствует ряду смежных означаемых (сочетанию фонем), согласно схеме: $\frac{\text{буква}}{\text{фонема} + \text{фонема}}$.

Во французском языке есть только одна такая графема буква *x*, обозначающая звукосочетание [ks]. Она восходит в конечном счете к букве (ξ) греческого алфавита, где была еще одна подобная графема ψ. Графема такого рода может быть названа с и н т е т и ч е с к о й, ибо она с помощью одного означающего передает то, что в других случаях изображается группой означающих. В самом деле, во многих французских словах [ks] изображается сочетанием cs (*tocsin*), *cc* (*accès*), *ct* (*action*) и др.

Орфографический с и н т е т и з м может быть сопоставлен с аналогичным явлением в других разделах языка [5]. Так во французской грамматике *moindre* является синтетическим знаком на фоне *plus petit*. Если бы последнего способа образования степени сравнения не существовало, то формы типа *moindre* были бы абсолютными, а не синтетическими знаками. Сопоставляя франц. *bru* и *gendre* с *belle-fille*, *beau-fils*, можно также

[3] Исключение составляют сложные слова и слова с префиксами: *vraisemblable*, *asymétrie* и т. д.

[4] Имеются немногочисленные исключения: *châsse*, *mess*, *que je vinsse*.

[5] См.: Ш. Б а л л и. Общая лингвистика и вопросы французского языка. М., 1955, тр. 164—169.

заметить, что первые два знака синтетические, а вторые два — аналитические (восходят к словосочетаниям).

2. Ряд смежных означающих (буквосочетание) соответствует одному означаемому (звуку): $\frac{\text{буква} + \text{буква}}{\text{фонема}}$.

Во французском языке таких графем очень много. Сюда относятся многочисленные буквосочетания, означающие один звук: *ph*, *ch*, *qu*, *ai*, *au*, *eau*, *œu* и др. Существует и другой тип буквосочетаний, которые обозначают не один, но два или даже три звука. Однако это не те звуки, которые данные буквы обозначают обычно в отдельности. Таковы немецкое *eu* [ɔɪ], а не [eu], французское *oi* [*wa*] и др. Этот тип может быть представлен так:

$$\frac{Б + Б_1}{Ф + Ф_1} \neq \frac{Б}{Ф} + \frac{Б_1}{Ф_1}$$

Оба типа таких составных буквосочетаний характеризуются своего рода идиоматичностью: их звуковое значение не является суммой значений входящих в них букв. Эта орфографическая идиоматичность может быть сопоставлена с лексической идиоматичностью, где также нарушено синтагматическое отношение между означаемым и означающими (ср. *tout-à-fait* и *complètement*) [6].

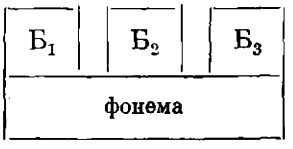
Такой подход может помочь решению вопроса о том, входит ли лигатура в графическую систему на правах отдельной графемы или же она представляет собой вариант последней.

Если лигатура имеет то же значение, что и сочетание входящих в нее букв, то она — вариант графемы. Так, во французских шрифтах сочетания *fi*, *fl*, *ct* и некоторые другие изображают иногда в виде лигатур. Они имеют такое же значение, как и соответствующие разъединенные знаки и должны рассматриваться как их варианты. Другое дело лигатура *œ*. Она имеет иное значение, нежели *oe*: Ср. *œstre* и *coexister*. Ее следует включить в число графем.

Выделение буквосочетаний в качестве графем является одним из сложнейших вопросов, и мы к нему вернемся ниже.

Помимо синтагматических, знаки находятся между собою в парадигматических отношениях. Если во всех случаях данной фонеме соответствует одна и та же буква и букве соответствует одна и та же фонема, то такой знак называется однозначным. Нарушение однозначности может быть двоякого рода.

3. Одна и та же фонема изображается разными буквами или буквосочетаниями

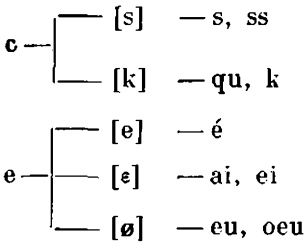


Так возникают синонимические знаки. Во французском языке такие случаи исключительно часты. Как и в других разделах языка, синонимия может быть полная и частичная. В первом случае оба знака всегда имеют одно и то же значение. Таковы *ô*, *eau*, *au*= [o], *ai*, *ei*, *ê*= =(ɛ), ç, ss= [s] и др. Во втором случае синонимия тесно связана с многозначностью. Одна и та же буква может иметь ряд синонимов в зависимости от ее конкретного значения.

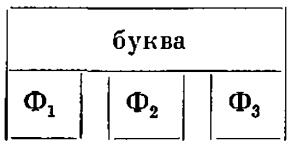
---

[6] Ш. Балли подобный знак называет фракционированным. См.: указ. соч., стр. 161—162.

Так



4. Одно означающее может быть использовано для нескольких означаемых. Так создаются полисемические графемы.



Подавляющее большинство букв и буквосочетаний французской графики многозначны.

Как и в других разделах языка, в области графики полисемический знак ставит две проблемы: а) отграничение его от знака-омонима; б) отграничение его от составного (идиоматического) знака.

С точки зрения структуры языка омонимами могут считаться только те знаки, у которых означающие имеют обязательные, отличные друг от друга варианты. Так, в лексике подлинными омонимами могут быть названы лишь так называемые частичные омонимы (например, *три* — число и *три* — от *тереть*), ибо сама языковая система здесь показывает, что это — формы разных слов. Напротив, «полные» омонимы (*ключ* 'родник' и *ключ* от замка) не имеют никаких структурных разграничительных признаков, и для того, чтобы отличить в этом случае омонимию от многозначности, приходится выходить за пределы собственно структуры языка — в область истории, этимологии. Не случайно именно в отношении «полных» омонимов возникает больше всего спорных и субъективных суждений.

В графике омонимами являются не знаки, но варианты знаков. Так, в немецком готическом шрифте совпадают заглавные I и J. Но тот факт, что эти буквы дифференцируются при написании в строке и в других разновидностях шрифтов, используемых немецким языком, заставляет видеть в них два омонимичных знака. В то же время различающиеся по форме итальянские *i* и *j* представляют в современном языке варианты одной графемы (ср. *operäi* и *operaj*).

Значительно труднее обстоит дело с отграничением многозначности от фразеологии. Для того, чтобы пояснить суть проблемы, обратимся к примерам из лексики. Глагол *prendre* в *prendre un livre* выступает в свободном значении. В *prendre corps* 'оформляться' он входит в состав неразложимого фразеологического сочетания. А как следует трактовать, например, сочетание *prendre le train* 'сесть в поезд'? Здесь можно видеть либо единое фразеологическое сочетание, либо сочетание глагола *prendre* в новом значении ('пользоваться транспортом') с соответствующим дополнением. Ведь вместо *train* можно употребить любое другое дополнение, обозначающее вид транспорта (*prendre le taxi*, *l'autobus*, *l'avion*, *le métro* и т. д.). Ограниченность круга дополнений определяется здесь лишь внеязыковыми факторами, реальной сочетаемостью идей и предметов. Та же проблема встает перед нами, когда мы рассматриваем многозначные буквы. Возьмем, к примеру, *e* в исходах *-er*, *-ez*. Поскольку конечные согласные во французском языке часто не читаются, мы можем сказать, что здесь

e имеет значение [e] перед согласными *r*, *z*. Но можно определить *-er*, *-ez* и как составные идиоматические буквосочетания со значением [e]. Здесь, следовательно, возможно двоякое объяснение: *-er=e*[e]+*r* (нуль) — многозначность *e* или *-er*=[e] — идиоматика *e*+*r*.

Признав *-er* буквосочетанием, мы должны также признать сочетанием *est* (в *il est*), *et*, *-er*, *-ez*, *-ed* (*pied*), *-ef* (*clef*), *-ect* (*aspect*).

Признав *-er* сочетанием полисемического знака с его дифференциатором, мы должны признать такую же природу и у всех перечисленных окончаний.

Итак, с точки зрения синтагматических и парадигматических связей могут быть 4 вида отступлений от а б с о л ю т н о г о (то есть отдельного и однозначного) знака во французской графике: синтетический, составной (идиоматический), синонимический и полисемический знаки.

Но этим еще не исчерпываются расхождения букв и звуков в графической системе. В алфавитной системе письма, где знак (буква) отражает прямо или косвенно звук, могут быть еще два вида расхождений, связанные с наличием означающего или означаемого. Синтагматические и парадигматические отношения создают условия для отсутствия означаемого или означающего в знаке.

5. Означающее может быть неполным, подразумеваемым [7]. Здесь образуется п о д р а з у м е в а е м ы й знак. В графике подразумеваемый знак можно обнаружить в сокращениях. Так *M.* вместо *Monsieur* есть типичный подразумеваемый знак. При чтении *M. Durand a dit* мы произнесем не [m], а полностью: *Monsieur*, однако в написании такой знак не имеет означающего. Получается схема.



Это опущение означающего может быть сопоставлено с эллипсом в области лексики и грамматики. Подразумеваемый знак не характерен для семиотической системы языка. Нарушенное соотношение между означаемым и означающим имеет тенденцию к восстановлению. Но это последнее происходит на базе сдвига, переосмысления первого или второго. Таким образом, здесь могут быть два случая:

1) О з н а ч а е м о е приводится в соответствие с означающим. Наиболее характерный пример этого в графике — абревиатуры. Когда мы сокращение читаем полностью (*SFIO* — «Section Française de l'Internationale Ouvrière»), перед нами опущение знака, устранение из написания обозначения звуков и слов. Если же мы то же сокращение читаем по буквам [*SFIO*], то происходит устранение эллипса и от подразумеваемого знака мы возвращаемся к полному, эксплицитному.

2) О з н а ч а ю щ е е переосмысляется. В связи с частым использованием оно становится особым синтетическим знаком для данного означаемого. Так *M.* обозначается всегда *Monsieur*, а не *Maître*, для которого используется знак *M*^{e}. В этом случае, оторвавшись от звуковой формы, знаки превращаются в символы понятий, живущие за пределами алфавита, подобно цифрам, знакам «+» и «—» и др. Когда мы *SFIO* читаем «*Le Parti Socialiste Français*», мы также символизируем этот знак.

С этим переосмыслением, ведущим к изменению природы знака, мы встречаемся и в лексике, и в грамматике. Так, *une station centrale* дало сокращенный знак с подразумеваемым компонентом *une centrale*; последнее, все шире употребляясь, стало самостоятельно обозначать то же понятие и теперь перед нами полный, хотя и полисемический знак — *centrale*.

---

[7] См.: Ш. Б а л л и. Указ. соч., стр. 175.

Латинское *Paulus aeger* могло рассматриваться на фоне более обыденного *Paulus est aeger* как эллипсис с подразумеваемым знаком. В русском языке эллипсис стал нормой и *Павел болен* — полный знак.

6. Означаемое отсутствует при наличии означающего: буква

В этом случае перед нами — пустой, нулевой, служебный знак. Исторически знак может утратить свое означаемое, которое лежало в основе его возникновения, перейдя при этом в иную категорию, в иную систему знаков. В графике такими являются нечитаемые буквы, весьма многочисленные во французском языке. Утратив собственное звуковое содержание, они приобрели служебную функцию и используются:

а) для создания устойчивых буквосочетаний: *h* в *ph*, *ch* и др;

б) для дифференциации значения многозначных букв. В этом их функция сходна с диакритическими знаками. Ср: *brochet* и *brocher*, где *t* указывает на открытое, а *r* на закрытое чтение [e]; *bas — base;*

в) для дифференциации слов-омонимов (ср: *doit* и *doigt*) и вообще для придания слову определенного индивидуального облика.

Как бы то ни было, немые буквы, утратив связь со звуком, из ряда алфавитных знаков переходят во внеалфавитные, дополнительные знаки.

Аналогичные явления наблюдаются и в других областях языка. В лексике самостоятельные слова могут превращаться в словообразовательные морфемы: нем. *Kaufmann*, фр. *géologue*. В грамматике самостоятельные формы слов превращаются в служебные элементы, в морфемы: *j'ai → j'ai dormi.*

Наличие нечитаемых букв и функционально родственных им диакритических знаков ставит трудную проблему при определении системы графем языка. Как отличить сочетание читаемой и нечитаемой буквы от слитного буквосочетания? Что такое *ph* — единое буквосочетание, единая графема или же две графемы, из которых первая — многозначная, а вторая — служебная, уточняющая значение первой. Что такое *th* — один знак, или два: *t*+*h*? Как определить, является ли диакритический знак отдельной графемой или же он составляет единое целое с буквой — новый знак?

Рассмотрим вначале проблему диакритических знаков. С точки зрения соотношения звука и буквы, дополнительные знаки, сопровождающие букву, неоднородны. Сравним, например, фр. *é* в *dé*, и испанское *é* в *diré* (будущее время от *decir*), фр. *père* и итальянское *tè*.

Во французских словах знаки *é* и *è* обозначают различные фонемы. Устранение диакритических знаков (accents) ведет к замене данной фонемы другой (*de*). В двух других языках надстрочные знаки служат либо для указания на место ударения в слове (*diré*) либо для различения омонимов (ит. *tè* 'чай' и *te* 'тебя'). Снятие знака изменяет общий звуковой облик или значение слова, но не отражается на его фонематическом составе (Исп. *e* и *é*, ит. *è* и *e* произносятся в приведенных выше словах одинаково). Следовательно, если французские знаки [´], [`] принадлежат букве, образуя с ней единую графему, то испанские и итальянские знаки принадлежат не букве, а слову. Только первые знаки могут называться диакритическими, т. е. различающими произношение букв. Но и сами диакритические знаки неоднородны. Французский знак [´] показывает закрытое произношение только у буквы *e*, у *o*, *eu* закрытость показывается иначе. Знак [ˆ] ставится над многими гласными буквами во французском языке, но показывает он всякий раз различную модификацию. Так *ê* означает открытое и долгое [ɛ:] в *tête*, открытое и краткое в *prêter*; *ô* изображает закрытое [o]; *î* — долгое [i] и т. д. За знаком [ˆ] во французской графике не закреплено никакое определенное фонематическое значение. Очень многие знаки используются как служебные модификаторы букв, употребляемые для создания нового знака, которого недостает в существующем алфавите.

Таково *ç* во французском языке, *ñ* в испанском *š*, *č*, *ž* в чешском и т. п. Нельзя сказать, что значок [ˇ] передает такую-то фонематическую модификацию основной фонемы.

Но в ряде случаев диакритический знак имеет собственное значение, и, присоединяясь к ряду букв, он повсюду сохраняет его. Таковы знаки долготы *ā*, *ē* в латинском, латышском и других языках, знак назализации в литовском *ų*, *į*, знак палатализации в польских *ń*, *ś*, *ć* и т. п. Первые диакритические знаки мы можем назвать идиоматическими (нерегулярными), а вторые — аналитическими (регулярными). Эти типы знаков можно сопоставить с непродуктивными и продуктивными аффиксами словообразования, с непродуктивными (фразеологизированными) и регулярными морфологическими изменениями. Итак, имеются три разновидности дополнительных знаков: идиоматические буквенные знаки, служебные буквенные знаки и иероглифические словесные знаки. Во французском языке преобладают идиоматические знаки. Аналитических диакритических знаков в нем нет. К словесным («иероглифическим») знакам могут быть отнесены частично accent grave в омонимах: *à*, *çà*, *où* (в отличие от *a*, *ça*, *ou*) и accent circonflexe в аналогичной функции: *crû*, *dû*, *je croîs* (в отличие от *cru*, *du*, *je crois*) и др.

Анализ функций диакритических знаков помогает решить вопрос о буквосочетаниях. Авторы трудов по французской орфографии склонны называть буквосочетанием любую комбинацию букв, чтение которой не соответствует чтению ее компонентов. Так называют буквосочетаниями и *eau*, *eu*, *ph*, *th* и *gt* в *vingt*, и *oa* в *toast* и т. д. Между тем здесь следует видеть различные факты.

а) Сочетание букв, как мы видели выше, может носить идиоматический характер: *eau* = [o]; *eu* = [œ], *ph* = [f] и т. п.

б) Могут быть и иные случаи, когда нечитаемая буква видоизменяет чтение основной соседней буквы, причем это видоизменение может иметь нерегулярный или регулярный характер. Так в немецком языке *e*, следующее за *i*, обозначает долготу, но то же *e*, следуя за буквами, обозначающими звуки заднего и среднего ряда, некогда показывало переход артикуляции вперед: *ae*, *oe*, *ue*. Таким образом в *ie e* выступает как единичный модификатор, а в архаических написаниях *ae*, *oe*, *ue* — как регулярный для данной группы букв модификатор. В первом случае *e* сливается с *i* в единое буквосочетание, образующее [*ie*]. Во втором *e* выступало как служебный диакритический знак. И действительно, оно легко заменяется аналитическим диакритическим знаком *oe* = *ö*, *ae* = *ä* *ue* = *ü*. Такие сочетания букв занимают промежуточное положение между буквосочетанием идиоматического типа и сочетанием двух букв — произносимой и немой. Нечитаемая буква выступает в качестве служебного знака и тогда, когда она обозначает модификацию, выражаемую внутри данной графической системы и с помощью диакритических знаков. Ср. *clef* и *clé*. Нечитаемая буква также выполняет служебную функцию, играя роль разделительного знака. Ср. франц. *u* в *guerre*; ит. *h* в *che* и т. п. Здесь она также может быть приравнена к диакритическому знаку ср. (*douceâtre* и *douçâtre*).

в) Наконец, нечитаемая буква, нисколько не влияющая на чтение окружающих букв, превращается в нулевой (по отношению к чтению данной буквы) знак и служит для оформления слова в целом, выполняя иероглифическую функцию.

Если соотносить написание со звучанием, нетрудно провести грань между буквосочетанием и сочетанием буквы (букв) с нечитаемой буквой. Так *ea* во фр. *speaker* — буквосочетание, в то время как *oa* в *toast*, *aou* в *saoul* — сочетание произносимых букв с немой буквой (*o* + *a*, *a* + *ou*).

Во французской графике представлены все три типа нечитаемых букв: компоненты неразложимых сочетаний, иероглифические нулевые буквы и (лишь в качестве отдельных элементов) встречаются служебные непроизносимые буквы [8].

Подчеркнем еще раз параллелизм служебных элементов графики непроизносимых букв и диакритических знаков:

	буква	знак
1. Компонент неразложимого сочетания (идиоматический знак)	польск. *sz*, *cz* в отличие от *s*, *c*	чешск. *š*, *č*
2. Служебный элемент		
а) единичный, нерегулярный модификатор	фр. *clef*	фр. *clé*
б) регулярный модификатор (аналитический)	нем. *ah*	лат. *ā*
в) разделительный знак	русск. *съезд*	русск. *с'езд*
3. Нулевой знак в составе слова (словесный иероглифический знак)	фр. *vingt*	фр. *dû*

История языков знает примеры, когда служебная (диакритическая) буква сливалась в единое сочетание с основной буквой или же превращалась в иероглифическую нулевую букву.

Придыхание, свойственное некоторым греческим согласным, в латинском языке передавалось буквой *h*. Так появились в латинском алфавите сочетания *ph*, *th*, *rh*, *ch*, где *h* выступал в качестве служебной буквы, имеющей специфическое значение в качестве регулярного модификатора. В разных европейских языках эти сочетания приобрели новую сущность. Они испытали либо тенденцию к сжатию, при котором сочетание превратилось в неразложимое единство, в буквосочетание, либо тенденцию к распаду, причем *h* перестало влиять на произношение предшествующей буквы и превратилось в пустой иероглифический знак.

Эта эволюция видна из схемы:

	Английский язык	Французский язык	Немецкий язык
ph	[f] — буквосочетание		
ch	[tʃ] буквосочетание [k] + разделительное *h* : *archives* [1] [k] + нулевое *h*: *character*	[ʃ] буквосочетание [k] + разделительное *h* : *orchestre* [1] [k] + нулевое h: *archaïque*	[x] буквосочетание
th	[θ] буквосочетание	[t] + нулевое *h*	
rh	[r] + нулевое *h*		

[1] *h* перед гласными переднего ряда может рассматриваться как нерегулярный разделительный знак, обозначающий, что *c* должно читаться как [к]. В этих случаях написание *h* мотивировано, в отличие от слов типа англ. *character*, фр. *archaïque*, где оно совершенно излишне. Таким образом, если в латинском языке были графемами все четыре сочетания, то в английском языке три (*ph*, *ch*, *th*), а во французском и немецком — только два (*ph*, *ch*).

Установление системы графем французского языка требует разрешения целого ряда проблем, на которых мы здесь не имеем возможности останавливаться; среди них — вопрос о двойных согласных буквах, о разгра-

[8] См.: В. Г. Гак. Французская орфография, изд. 2-е. М., 1959, стр. 9.

ничении переплетающихся буквосочетаний в слове, о тенденции к превращению в графемы сочетаний *muta cum liquida*: *cr*, *tr*, *bl*, *cl*, дистрибуция которых во многом совпадает с дистрибуцией простых согласных графем, и другие.

Если многозначные графемы делить на ряд отдельных субграфем, различающихся по их означаемым, то во французском языке существует несколько десятков таких субграфем: *a* [ɑ], *a* [a], *à* [ɑ], *â* [ɑ], *â* [a], *a* [—], *ai* [ɛ], *ai* [e], *b* [b], *b* [—], *c* [k], *c* [s], *c* [—] и т. д.

МЕТОДЫ СТРУКТУРНОЙ ЛИНГВИСТИКИ И ИХ ПРИМЕНЕНИЕ К ИЗУЧЕНИЮ ОРФОГРАФИИ

Если графика определяет средства, которыми располагает алфавит данного языка для изображения его фонем, то орфография изучает факторы, влияющие на выбор при написании той или иной графемы. Орфографический анализ должен установить отношения между графемами. Графемы, подобно другим языковым знакам, связаны, как это определил Соссюр, отношениями двоякого порядка: синтагматическими и парадигматическими.

С синтагматическими отношениями языковых единиц непосредственно связан дистрибутивный анализ [9]. Он показывает сочетаемость языковых единиц (в нашем случае — графем) и дает возможность установить дистрибутивные структуры — значимые (с точки зрения данного уровня анализа языка) отрезки речевой цепи.

Парадигматические отношения исследуются парадигматическим анализом. Выделяемые при этом системы характеризуются некоторой общностью (значения, структуры, происхождения и др.), не вытекающей непосредственно из одной лишь сочетаемости единиц. Парадигматический анализ позволяет установить наличие многозначности языковой единицы [10].

Существуют два основных вида языковой деятельности: восприятие речи и ее произведение (эмиссия). В приложении к письменной форме языка это означает чтение и письмо. Орфография (письмо) неразрывно связана с правилами чтения. Поэтому мы дистрибутивный и парадигматический анализ приложим к обеим сторонам речевой деятельности: чтению и письму.

Чтение. При правилах чтения ярко проявляются синтагматические отношения знаков, анализ которых позволяет определить дистрибутивные структуры. В условиях многозначности, служебности и других отклонений от абсолютного знака, свойственных французским графемам (см. выше), определение дистрибутивных структур имеет особенно важное значение. При дистрибутивном анализе учитывается последующее, предшествующее и промежуточное положение основного элемента (X+a, a+X, a+X+b). Французской графике свойствен позиционный принцип чтения букв и, в частности, связь графем с последующим знаком. Буква или буквосочетание получает точное значение лишь по связи с другими буквами, особенно последующими [11] (тип: X+a). Разные дистрибутивные структуры могут включать в свой состав разные ядра при одних и тех же вспомогательных элементах. Так,

---

[9] См.: С. К. Шаумян. Лингвистические проблемы кибернетики и структурная лингвистика. «Вопросы философии», 1960, № 9; его же: О сущности структурной лингвистики. «Вопросы языкознания», 1956, № 5.

[10] Такой же является, как известно, одна из функций трансформационного анализа. Однако, во избежание терминологической путаницы, мы воздержимся от отождествления отмеченного выше парадигматического анализа с трансформационным, имеющим более широкое содержание.

[11] См.: В. Г. Гак. Указ. соч., стр. 11.

структуры: «ядро+гласная буква или *m*, *n*», «ядро+согласная буква или ноль» выделяются для буквосочетаний: «гласная+*m* или *n*».

Например: *an*, *rang* [ã], *cane* [an]
on, *pont* [ɔ̃], *zone* [on]
en, *dent* [ɛ̃], *venir* [ən]

Структуры: «ядро+гласная переднего ряда (*i*, *e*, *y*)», «ядро+гласная заднего ряда, или согласная, или ноль» выделяются для букв *c*, *g*.

Например: *cécité* [s], *cas* [k]
geler [ʒ], *gant* [g]

Структуры: «ядро+произносимая согласная, кроме *s*», «ядро+непроизносимая согласная, или *s*, или ноль» выделяются для *o*, *eu*, *oeu*.

Например: *porter* [ɔ], *rose*, *pot*, *vélo* [o]
beurre [œ], *creuse*, *feux*, *peu* [ø]
soeur [œ], *noeud*, *voeu* [ø]

Структуры: «ядро+согласная или немая гласная», «ядро+произносимая гласная» характерны для *ou*, *u*, *i*.

Например: *loup* [u], *louer* [w]
mur [y], *muet* [ɥ]
vie [i], *hier* [j]

Во французской графической системе имеются также структуры, построенные по типу a+X+b; a+X.

Примером структуры с двусторонней связью ядра (a+X+b) могут служить: «гласная+*s*+гласная»; «гласная+*il*+ноль»; «гласная+*ill*++гласная»; «*c* (или *g*)+*ue*+*il*» и др.

Структуры с регрессивной связью (a+X) во французской графике очень редки.

Но дистрибутивные структуры не исчерпывают всех правил чтения французского языка. Определенное значение, хотя и меньшее, имеют в них и парадигматические структуры. Поясним примером.

Выше мы упоминали дистрибутивные структуры «*en* + гласная», «*en* +*n*», которые, противопоставляясь структуре «*en* + согласная или ноль», дают соответственное чтение [ən] или [ɛn]. Однако в слове *enivrer en* произносится [ã], а не [ən], в слове *ennoblir* = [ã], а не [ɛn]. Причина этого кроется в том, что в этих словах *en* — префикс с определенным значением (в данном случае — придание некоторого качества). Его чтение зависит уже не от синтагматических связяй с последующей буквой, как чтение *-en-* в словах *tenir* или *renne*, но от парадигматических связей (выделение в группе слов определенного общего значимого элемента).

Таким образом, *en* в начале слова образует особого рода структуру не синтагматического, но парадигматического характера. Поэтому, встретив *en* в начале слова (*enneigé*, *ennemi*), мы можем его толковать двояко: [ã] или [ɛn]. Дистрибутивный анализ здесь недостаточен. Необходимо сопоставить эти слова с другими элементами языка, определить, является ли или нет отношение *en-*+*-neigé* видоизменением отношения, лежащего в основе связи элементов в *en-*+*-nobli*. В данном случае ответ положительный, и мы должны читать *en* как [ã]. Отношение *en-*+ +*-nemi* в *ennemi* не может быть трансформировано в отношение равнозначное *en-*+*-nobli*, и, значит, *en* не является префиксом и должно читаться [ɛn].

Дистрибутивный анализ определяет, что *ent* должно читаться [ã]: *dent*, *lent*. Однако в глагольных окончаниях *-ent* не произносится (*ils parlent*). Поэтому встретив форму типа *président*, нужно, путем сопоставления с другими формами языка, определить, является ли она глаголом или именем (*ils président — ils parlent; un président — un parent*), и в зависимости от этого решать вопрос об истинном звуковом значении буквенных элементов. И здесь также парадигматические отношения ограничивают сферу деятельности синтагматических, и парадигматический анализ приходит на помощь дистрибутивному. Выше мы видели два типа парадигматических структур: словообразовательную (*ennoblir*) и морфологическую (*ils président*). К ним можно добавить этимологические структуры (так, в словах латинского происхождения *en* произносится [ɛ̃] — *examen* или [ɛn] *abdomen*), и единичные (*solennel*), распространявшиеся на слова данного гнезда, на отдельные единицы.

Письмо. В выборе графем при написании также действуют синтагматические или парадигматические отношения, в связи с чем слова подвергаются дистрибутивному и парадигматическому анализу.

Дистрибутивный анализ выявляет сочетаемость графем. Сочетаемость эта имеет двоякое основание. Во-первых, она зависит от реальной сочетаемости звуков в языке. Во-вторых, она зависит уже от закономерностей системы орфографии. Так, во французском языке можно встретить дистрибутивную структуру *str* (*structure*), но почти нельзя обнаружить *stk* или *stv*, распространенную в русском языке (*повестка*, *ствол*), только потому, что сами последние звукосочетания несвойственны французскому языку. Но то обстоятельство, что во французских словах встречается сочетание *gi* и почти не встречается *ji*, объясняется тем, что для обозначения одного и того же звукосочетания [ʒi] используется первый знак, а не второй. Наличие дистрибутивной структуры *gi* и отсутствие *ji* есть факт уже чисто внутриорфографический. С аналогичным явлением мы сталкиваемся, например, и в синтаксисе словосочетаний. Сочетаемость слов порой определяется факторами, лежащими вне синтаксико-фразеологической системы (сочетаемость самих понятий), а иногда выступает как проявление внутренних закономерностей синтаксиса.

Во французской орфографии дистрибуция имеет огромное значение. Подобно правилам чтения, орфография проявляет тяготение к позиционному принципу, т. е. к зависимости выбора написания от соседних букв. В орфографии наблюдаются как прогрессивные структуры (X+a), так и регрессивные (a+X). Примером прогрессивных структур (X+a) является написание accent circonflexe в зависимости от последующей согласной, неудвоение согласной в зависимости от последующих букв (*ferré*, но *fer*) и др. Но преобладающими структурами в орфографии являются регрессивные, в то время как в правилах чтения — прогрессивные (орфография и правила чтения как бы зеркально отражают друг друга). Примером структур a+X является написание двойных согласных в зависимости от предшествующей гласной, так, например, *n* после *a*, *o*, *e* удваивается, после других букв — нет (т. е. возможны структуры *a+n*, *a+nn*, *o+n*, *o+nn*, но только *ai+n*, *ou+n*, *u+n* и т. д.). Написание звука [j] также зависит от предшествующей гласной, написание *e* muet на конце слова — от последней произносимой согласной и т. п. Дистрибуция определяет выбор между *au* и *eau* и многие другие явления [12].

Но как бы ни была важна дистрибуция для французской орфографии, она не дает полного анализа функционирования всех единиц. Очень часто приходится прибегать к парадигматическому анализу.

---

[12] См.: В. Г. Гак. Указ. соч., соответствующие параграфы.

Способ написания звука [ɛ]: *è, e, ê, ai, ei* в значительной степени определяется дистрибуцией. Так перед двумя согласными может быть только *e* или *ai* (перед *ss*): *presser, baisser*. Написания *ê, è, ei* здесь исключаются. Но как определить, следует ли писать в *presser, baisser e* или *ai?* В этом случае дистрибуция нам ответа дать не может, и становятся важными парадигматические отношения, т. е. связь с другими словами корня. Слово *presser* мы должны рассматривать как трансформацию слов *près, presse* и т. п. Слово *baisser* следует трактовать как трансформацию однокоренных *bas, basse*. Эта трансформация показывает нам, что в корне *press-* [ɛ] никогда не чередуется с [а], а в *baisser* есть такое чередование. Поэтому в первом случае мы используем графему *e*, а во втором *ai*, которая обозначает звук [е] при чередованиях *a/e*. Ср. *pain — panier, clair — clarté, prolétaire — prolétariat* и т. п.

При парадигматическом анализе учитываются различные виды ассоциаций: морфологические, словообразовательные, этимологические и др.

ОПРЕДЕЛЕНИЕ ПОНЯТИЯ ОРФОГРАФИЧЕСКОГО ПРАВИЛА

На основании вышеизложенного мы можем теперь дать определение орфографического правила, исходя из структурного метода.

Орфографическое правило есть обоснование выбора написания с точки зрения дистрибутивного или парадигматического анализа. По сути дела, все существующие орфографические правила во всех языках, пользующихся алфавитной системой письма, сводятся либо к правилам дистрибуции, либо к правилам трансформации.

Так, когда мы говорим, что в русских словах после шипящих не пишутся *ы, э, ю, я*, то мы тем самым формулируем правило дистрибуции, сочетаемости букв. При этом мы игнорируем различие в том, как эта дистрибуция соотносится с объективным произношением. Так, написания *шу, ша* соответствует реальному произношению ([ш] в русском языке — твердая фонема), в то время как *ши, ше* противоречат звучанию.

Когда мы формулируем правило написания безударных гласных или конечных согласных в русском языке, мы тем самым прибегаем к парадигматическому анализу, ибо рекомендуем трансформировать полисемические единицы таким образом, чтобы их сущность обнажилась (ср. *вода — вóды, зуб — зубы*).

При общих основах оба вида структур проявляют большое разнообразие по языкам и создают национальную специфику орфографий, даже при общности алфавитов.

Так, французское письмо отличается от немецкого или английского не только наличием значков *è, é, ê* и др., но, пожалуй, в еще большей степени дистрибуцией одинаковых графем. Например, во французском языке сочетание *dt*, характерное для немецкого (*Stadt, Schmidt*), настолько немыслимо, что даже в глаголах типа *rendre* его не пишут в 3 лице: *il rend* вм. *il rendt*. В нем не встречаются сочетания типа *oo, ow, aw* и др., типичные для английского языка, написание двойных согласных на конце слова и т. д.

Парадигматические системы также значительно различаются в разных языках. Трансформации в пределах одного корня (морфологический принцип) характерны для многих орфографий. Но содержание их различно. Так, в русском и французском языках прибегают к трансформации для определения написания конца слова. Но при этом в русском языке выявляется соотношение «глухой согласный»: «звонкий согласный», а во французском языке «нуль звука»: «согласный» (ср. *год — годы* и *rond — ronde*). В обоих языках используют трансформацию для определения способа обозначения гласного звука. Но если в русском языке

при этом устанавливается система «безударный»: «ударный гласный», то во французской системе: *ai|a, ei|e, au|al, eau|el, ê|es* и др.

Внедрение методов структурной лингвистики дает возможность научно описать орфографию языка в синхронном плане. Особенно это важно для такого языка, как французский. В течение десятилетий люди, изучавшие французскую орфографию, видели в ней почти исключительно этимологический принцип. Любое написание они старались вывести из латинского языка, игнорируя те соотношения между графемами, которые имеются в современном языке. Тщательное статистическое изучение сочетаемости графем, равно как и анализ написаний с точки зрения их соотношения с лексико-морфологической структурой языка дает возможность подвести научное обоснование под существующие правила чтения и письма, а также обнаружить ряд новых правил.

ORTHOGRAPHY IN THE LIGHT OF STRUCTURAL ANALYSIS

Summary

In the present paper the author applies the methods of structural linguistics to analysing the orthography of a language (the language in question being French).

The graphic system of any language consists of some fundamental units — graphemes — which can have positional, functional, individual and national variants.

The absolute correlation between *le signifiant* and *le signifié* is often affected by syntagmatic and paradigmatic relations peculiar to the language system as a whole. Thus, there are six kinds of non-absolute sign-graphemes: synthetic, idiomatic, synonymic, polysemantic, elliptic and an auxiliary one. Similar types of signs can be found in other language subsystems (morphology, syntax, phraseology, vocabulary, word-building, etc.). To determine the criteria for them is one of the most important tasks of linguistics. The present paper deals with such criteria in orthography.

Structural linguistics offers two fundamental methods for defining most effectively the spelling rules. These are distributional and paradigmatic (resp. transformational) methods based on syntagmatical and paradigmatical relations respectively. Those two methods make it possible to bring out certain new laws in French orthography.

To give an orthographical rule is to explain the choice of a grapheme from the point of view of the distributional or paradigmatic analysis.

СОДЕРЖАНИЕ

Предисловие 5

I. ОБЩИЕ ПРОБЛЕМЫ

С. К. Шаумян. Преобразование информации в процессе познания и двухступенчатая теория структурной лингвистики 5
S. K. Šaumjan. Processing of information in cognition and the two level theory of structural linguistics (Summary) 12
И. И. Ревзин. О некоторых вопросах дистрибутивного анализа и его дальнейшей формализации 13
I. I. Revzin. On some questions of distributional analysis and its further formalization (Summary) 21
С. Маркус (Бухарест). Теория графов, лингвистические оппозиции и инвариантная структура 22
S. Marcus. Graph theory, linguistic oppositions and invariant structure (Summary) 30
Ю. К. Лекомцев. К вопросу об аналогиях в строении схем слога и простого предложения. 31
Y. K. Lekomtsev. On the similarities in structure between syllable and sentence (Summary) 42
Р. Г. Пиотровский. О теоретико-информационных параметрах устной и письменной форм языка. 43
R. G. Piotrovsky. On theoretico-informational parametres of the oral and written forms of language (Summary) 59
В. Ю. Розенцвейг и Л. М. Уман. К проблеме грамматической интерференции. 60
W. Y. Rosenzweig and L. M. Uman. On the problem of grammatical interference (Summary) 73

II. ФОНОЛОГИЯ

С. К. Шаумян. Панхроническая система дифференциальных элементов и двухступенчатая теория фонологии 75
S. K. Šaumjan. Panchronistic system of distinctive features (Summary) . . 79
И. И. Ревзин. Об одном подходе к моделям дистрибутивного фонологического анализа. 80
I. I. Revzin. An approach to the models of distributional phonological analysis (Summary) 85
Э. Василиу (Бухарест). Фонологическое описание румынского вокализма 86
E. Vasiliu. A phonological description of the Rumanian vocalism (Summary) 91
Р. Г. Пиотровский. Сосуществующие фонетические системы и стилистические корреляции в молдавском языке. 92
R. G. Piotrovsky. Coexistent phonemic system and stylistic correlations in Moldavian (Summary) 99

III. СТРУКТУРНАЯ ГРАММАТИКА

С. Я. Фитиалов. О моделировании синтаксиса в структурной лингвистике 100
S. Y. Fitialov. On syntax modelling in structural linguistics (Summary) 114
К. И. Бабицкий. К вопросу о моделировании структуры простого предложения. 115
K. I. Babitsky. On modelling the structure of the simple sentence (Summary) 129
Л. Н. Засорина. Модель именного склонения для русского письменного текста. 130
L. N. Zasorina. A model of the noun declension for written Russian (Summary). 140

IV. СТРУКТУРНАЯ ЛЕКСИКОЛОГИЯ

Ю. Д. Апресян. О понятиях и методах структурной лексикологии . 141
Y. D. Apresyan. On the notions and methods of structural lexicology (Summary) 162
В. И. Перебейнос. К вопросу об использовании структурных методов в лексикологии. 163
V. I. Perebejnos. On structural methods in lexicology (Summary) 174
П. А. Соболева. Компонентный анализ значений глагола на основе словообразовательного признака 175
P. A. Soboleva. Componental analysis of verbal meanings on the basis of word building properties (Summary) 189
Ю. К. Лекомцев. К вопросу о системности глаголов речи в английском языке 190
Y. K. Lekomtsev. On the pattern of verba dicendi in English (Summary) 197

V. ГРАФЕМИКА И СТРУКТУРНАЯ ТЕОРИЯ ОРФОГРАФИИ

Б. В. Сухотин. Экспериментальное выделение классов букв с помощью электронной вычислительной машины. 198
B. V. Sukhotin. Experiment in classifying letters by computor (Summary) 205
В. Г. Гак. Орфография в свете структурного анализа 206
V. G. Gak. Orthography in the light of structural analysis (Summary) 221

Проблемы структурной лингвистики

•

Утверждено к печати Институтом русского языка

•

Редактор издательства *В. А. Ицкович*
Технический редактор *Г. Н. Шевченко*
Корректор *В. Т. Макаров*

•

Сдано в набор 13/IV 1962 г. Подписано в печать 22/IX 1962 г.
Формат 70×108¹/₁₆. Печ. л. 14+1 вкл.
Усл. п. л. 19,69 (19,18+0,51).
Уч.-изд. л. 18,4 (18,1+0,3 вкл.). Тираж 4000 экз.
Т-11328. Изд. № 547. Тип. зак. № 656.

Цена 1 руб. 20 коп.

Издательство АН СССР. Москва Б-62, Подсосенский пер., 21
1-я тип. Издат. АН СССР. Ленинград, В-34, 9 лин., д. 12.

ОПЕЧАТКИ

Страница	Строка	Напечатано	Должно быть
67	табл. 4 раздел б) 2-я колонка справа	c i	Passé composé, imparfait
130	19 св.	естствен-ных	естественных

Цена 1 р. 20 к.