

Ю. Д. АПРЕСЯН

ИДЕИ И МЕТОДЫ
СОВРЕМЕННОЙ
СТРУКТУРНОЙ
ЛИНГВИСТИКИ

(краткий очерк)

574100

ИЗДАТЕЛЬСТВО «ПРОСВЕЩЕНИЕ»

Москва 1966

ПРЕДИСЛОВИЕ

Существует следующее предание о халифе Омаре, арабском завоевателе XII века. В одном из покоренных Омаром городов было найдено великое множество книг. Полководец спросил его, следует ли разделить книги, вместе с другой добычей, среди правоверных. Омар ответил: «Если в этих книгах говорится то, что есть в коране, они бесполезны. Если же в них говорится что-нибудь другое, они вредны. Поэтому и в том и в другом случае их надо сжечь».

Мы вспомнили об этом потому, что в течение ряда лет к структурной лингвистике относились приблизительно так же, как халиф Омар — к найденной им библиотеке. Новые идеи структурной лингвистики были объявлены вредными, а все остальные положения — давно известными истинами, облаченными в другой терминологический наряд.

Из числа новых идей наиболее ожесточенной критике подверглось введенное еще Фердинандом де Соссюром представление о языке как «системе чистых отношений» и восходящее к нему же представление о лингвистике как формальной теории, изучающей объекты, существование которых не выводится непосредственно из наблюдаемых языковых фактов.

В дальнейшем будут подробно обсуждены эти идеи Ф. де Соссюра и будет показано, что за ними не скрывается ничего, кроме попытки найти простые и общие законы, лежащие в основе всех языковых явлений. Здесь же ограничимся указанием на следующий любопытный факт: если бы Ф. де Соссюр в своей исследовательской работе не вышел за пределы обычных представлений и методов современной ему лингвистики и не постулировал языковых форм, существование которых непосредственно

не подтверждалось данными известных в его время языков, он не сделал бы одного из величайших в истории лингвистики открытий — открытия ларингальных (существо этого открытия излагается на стр. 94—95). Наличие ларингальных в определенном классе индоевропейских корней, предсказанное им на основании чисто теоретических соображений о структуре корня, было фактически подтверждено уже после его смерти, когда был открыт и дешифрован хеттский язык.

Перейдем ко второму пункту критики. В этом случае, как и в предыдущем, ограничимся ссылкой на прецедент, однако, несколько иного рода. Известно, что синтаксическая теория А. Х. Востокова включала учение о размещении слов в предложении. Это учение состояло из 1) правил порядка главных частей предложения (групп подлежащего и сказуемого), 2) правил размещения слов определительных и дополнительных внутри этих групп <32>¹. Сходные идеи используются в современных синтаксических алгоритмах автоматического синтеза текста <13>, <64>, <124>. Если бы эти алгоритмы не содержали, помимо терминов, ничего нового, они столь же мало допускали бы реализацию в виде программы для электронной вычислительной машины, как и правила А. Х. Востокова. Следовательно, новым в них является уровень точности языка, на котором они изложены.

О том, какое исключительное значение имеет точность языка науки, прекрасно сказано в очень глубоком афоризме Л. Витгенштейна, заключающем его «Логико-философский трактат» <38>: «О чем невозможно говорить, о том следует молчать». Наши интуитивные знания об окружающей нас действительности и особенно о нас самих практически безграничны, но на языке точной науки к настоящему времени могла быть записана лишь ничтожная их часть. В сущности, спор между структуралистами и неструктуралистами сводится к вопросу о том, может ли лингвистика стать точной наукой, или природа ее объекта такова, что она обречена всегда оставаться гуманитарной дисциплиной.

¹ Цифра в угловых скобках, данная прямым шрифтом, обозначает номер названия работы из списка литературы, помещенного в конце книги. Если в тексте приводится цитата, то страница цитируемого издания дается курсивом после номера работы. Порядок ссылок в большинстве случаев хронологический.

Соссюровское понимание языка и языкознания с вытекающими из него новыми задачами и неразрывно связанное с ним стремление перевести накопленные в лингвистике знания на язык точной науки исчерпывают существо структурного подхода к языку. Автор надеется, что текст этой книги даст читателю возможность составить независимое суждение о том, насколько такой подход оправдан.

Теперь нам остается сделать второе замечание, касающееся предлагаемой вниманию читателя книги. Эта книга не является ни систематическим курсом структурной лингвистики, ни хотя бы введением в нее. Это — очерк, цель которого состоит в том, чтобы ввести читателя в проблематику и методы современной структурной лингвистики и подготовить его к чтению специальной литературы. Форма очерка освободила автора от необходимости сказать понемногу обо всем и позволила сосредоточиться на проблемах, в основном, из области морфологии, синтаксиса и семантики, которые по той или иной причине казались ему наиболее заслуживающими внимания.

Возможно, что не все исследователи, занимающиеся структурной лингвистикой, согласятся с тем пониманием этой научной дисциплины, которого придерживается автор. Возражения может вызвать как выбор тем, так и выбор имен, которыми структурная лингвистика представлена в данной книге. Хотя такая ситуация более или менее неизбежна в период бурного развития молодых наук, автор хотел бы подчеркнуть, что он умышленно выбрал наиболее широкое из всех возможных толкований термина «структурная лингвистика», так как именно такое понимание предмета позволило ему включить в книгу и очерк классических школ лингвистики, известных под названием структурных, и обзор новых направлений, развивающих их идеи.

Книга построена по следующему плану. В части I дается очерк истории классической структурной лингвистики и ее школ. Во части II рассматриваются основные вопросы, связанные с понятием лингвистической модели — центральным понятием современной структурной лингвистики. В трех последних частях излагаются различные типы лингвистических моделей.

В заключение каждого самостоятельного раздела рассматривается обычно то или иное конкретное исследование, которое иллюстрирует применение описанного в данном разделе метода к решению определенной лингвистической задачи. Читателя не должно смущать то обстоятельство, что здесь наряду с работами признанных авторитетов рассматриваются исследования молодых ученых. По глубокому убеждению автора, некоторые из этих ученых получили значительные результаты.

Книга популярна в том смысле, что не требует от читателя никакой специальной подготовки помимо знания курса «Введение в языкознание», например, в объеме образцовой книги А. А. Реформатского <155>. Однако от неподготовленного читателя она требует некоторого упорства и прилежания. Только внимательное чтение книги способно ввести его в изучение специальной литературы по современной структурной лингвистике.

После того как работа над рукописью была завершена, в разных издательствах вышли в свет посвященные структурной лингвистике и смежным с ней дисциплинам исследования И. А. Мельчука, Б. А. Успенского, С. К. Шаумяна, З. М. Волоцкой, Т. Н. Молошной и Т. М. Николаевой, Н. Хомского, Дж. Катца и Дж. Фодора, И. Вахека и др. Кроме того, И. И. Ревзиным, А. А. Зализняком, А. В. Гладким, Е. В. Падучевой были защищены диссертации, имеющие непосредственное отношение к теме нашей книги. К сожалению, материалы указанных исследований не могли быть учтены в тексте этой работы в полной мере.

Автор пользуется случаем, чтобы выразить искреннюю признательность А. Вержбицкой, Е. А. Земской, Л. Н. Иорданской, П. С. Кузнецову, И. А. Мельчуку, Б. В. Сухотину и Р. М. Фрумкиной, которые прочитали книгу в рукописи и сделали ряд исключительно ценных критических замечаний.

ИЗ ИСТОРИИ СТРУКТУРНОЙ ЛИНГВИСТИКИ

Глава I

ПОЧЕМУ ВОЗНИКЛА СТРУКТУРНАЯ ЛИНГВИСТИКА

Новая научная дисциплина обычно возникает и развивается под влиянием внешних и внутренних стимулов.

В числе внешних стимулов, способствовавших развитию структурной лингвистики в последнее время, можно назвать некоторые крупные технические достижения последних десятилетий XX века. Подобно тому, как изобретение электронного микроскопа ознаменовало новый — молекулярно-синтетический — этап в развитии биологии, создание быстродействующих электронных вычислительных машин стимулировало развитие лингвистики. Эффект этого технического достижения был двойным: 1) возникло так называемое информационное дело (машинный перевод, автоматическое реферирование текстов, поиск информации и т. п.), поставившее перед лингвистикой новые требования; 2) появилась возможность механизировать трудоемкие, но не требующие значительных творческих усилий лингвистические работы, открывшая перед лингвистикой новые перспективы. Машины позволяют ставить крупномасштабные задачи, связанные со сплошным обследованием больших массивов текста, решение которых старыми средствами было бы невозможно. Интересно, что за рубежом (в Англии, США, Франции, ФРГ) работы по составлению частотных и обратных словарей, различного рода конкордансов и т. п. в последнее десятилетие практически переданы машинам <192>. Машины выполняют иногда и несложные лингвистические исследования, связанные, например, с выделением наиболее распространенных синтаксических структур некоторого языка и подсчетом их частотности <114>.

Однако для того чтобы воспринимать, хранить, выдавать или обрабатывать информацию, например выпол-

нять перевод текстов, машина должна знать язык, на котором эта информация записана. Чтобы исследовать текст, машина также должна иметь некоторые сведения о языке (например, чтобы выделить синтаксические структуры языка, машина должна уметь различать классы слов или члены предложения). Если машина не знает языка, то должна знать по крайней мере правила обработки текста, с помощью которых она могла бы извлечь из текста некоторые сведения об устройстве языка, на котором этот текст написан, и т. д. Во всех этих случаях человек должен вступить в общение с машиной и преподать необходимые машине знания в понятном для нее виде, т. е. формально.

Когда реально возникла задача обучения машины языку, оказалось, что обычная описательная грамматика с этой задачей справиться не в состоянии, потому что принятые в ней лингвистические описания не являются формальными. Возникла необходимость создать более точные описания языка, понятные не только человеку, но и современным вычислительным машинам.

Однако как бы велико ни было значение этих внешних факторов, роль в н у т р е н н и х факторов в становлении науки намного важнее. Исторически структурная лингвистика возникла гораздо р а н ь ш е, чем стимулировавшие ее развитие вычислительная техника и те или иные практические потребности. Она была реакцией на обычную описательную грамматику.

Описательная грамматика накопила богатейший фактический материал и большой опыт исследования связей между различными языковыми категориями. Однако она не создала точных понятий о языковых объектах, и поэтому имеющиеся в ней утверждения не поддаются проверке. Понятия описательной грамматики вырабатывались без общего плана. На протяжении тысячелетий каждая эпоха делала свой вклад в создание лингвистической терминологии, но основные термины никем не были сведены в единую, последовательно построенную систему. Это дало Ф. де Соссюру основание заявить, что в лингвистике мы оперируем единицами, которые как следует не определены, а А. Мейе еще более резко заметил по тому же поводу, что в лингвистике имеется столько же разных лингвистик, сколько есть лингвистов. В этих условиях крупнейшие языковеды предпринимают с конца

XIX века критический и конструктивный пересмотр оснований описательной грамматики, в результате которого и возникла структурная лингвистика.

Внимание было обращено прежде всего на неточность основных понятий, которыми оперирует традиционная описательная грамматика¹. В качестве примера мы кратко рассмотрим понятие «слово».

Наиболее распространенным, хотя и не общепринятым, является предложенное А. Мейе и включенное в «Словарь лингвистических терминов» Ж. Марузо <107> определение слова как «результата сочетания определенного значения с совокупностью определенных звуков, пригодного для определенного грамматического употребления». Допустим сначала, что нам уже известно, что такое «значение», «совокупность звуков» и «грамматическое употребление». Тогда мы обнаружим, что определение А. Мейе, помимо множества слов, применимо по крайней мере к множеству словосочетаний и множеству морфем (в том числе, грамматических). Действительно, у любого словосочетания, корня и аффикса имеется «определенное значение» и «определенное звучание», и все они способны к «определенному грамматическому употреблению». Если, кроме того, допустить, что предложению тоже свойственно «определенное грамматическое употребление» (ниже мы покажем, что такое допущение имеет смысл), то и предложение должно считаться словом; в противном случае мы рискуем оказаться непоследовательными.

Другим недостатком этого определения является то обстоятельство, что в нем никак не учтен факт расхождения в языке фонетических показаний с грамматическими и семантическими и грамматических с семантическими. Так, во французском языке благодаря наличию в нем ритмического ударения слова не имеют фонетических границ; поэтому фонетическое членение речевого потока не совпадает с грамматическим и семантическим, ср. *Pierre / a / besoin / de / ce / livre* — «Пьеру нужна эта книга»

¹ Читателя, интересующегося деталями этой критики, мы отсылаем к работам В. В. Виноградова <30>, <33>, <35>, П. С. Кузнецова <86>, М. В. Панова <135>, А. И. Смирницкого <166>, Л. В. Щербы <224>, <226> и др. Для студентов особенно полезной будет лаконичная, но очень содержательная брошюра П. С. Кузнецова <86>.

(показаны грамматические границы между словами) и [ˈpʲɛ:r/ a-bə-ˈzw̩ / də-sə-ˈli:vɐ] (показаны фонетические границы между ритмическими группами). Регулярным примером расхождения грамматических и семантических показаний являются аналитические формы (ср. *будет работать*, *will work*, *wird arbeiten* и т. п.) и фразеологические единицы (ср. *дать нагоняй* — «выбранить»): и те и другие иногда определяются как «образования», представляющие собой одно слово с семантической точки зрения и два или более слова — с грамматической. В результате единое понятие слова расщепляется на такие понятия, как «фонетическое слово», «лексическое слово», «грамматическое слово», которые никак не могут его заменить <30>.

Теперь рассмотрим термины «совокупность звуков», «грамматическое употребление» и «значение». В определении А. Мейе они фигурируют как независимые друг от друга: из структуры определения следует, что ни один из них не является производным от двух других терминов. Посмотрим, действительно ли мы можем оперировать термином «совокупность звуков», не зная хотя бы грамматических форм слова. Термин «совокупность звуков» в указанных условиях может значить только, что у каждого слова имеется более или менее постоянный звуковой состав, все возможные изменения которого могут быть объяснены как результат особых фонетических условий (например, в русском языке оглушения звонких перед паузой, озвончения глухих перед звонкими и т. п.). Достаточно, однако, вспомнить любую грамматическую парадигму, особенно парадигму с морфологическими (необъяснимыми фонетически) чередованиями типа *друг — друзья*, *бегу — бежал*, супплетивными формами типа *идет — шел*, *человек — люди*, *хороший — лучшие* и синтетическими и аналитическими формами типа *читает — читал — будет читать*, чтобы понять ошибочность нашего предположения о независимости термина «совокупность звуков» от термина «грамматическая форма». Очевидно, под «совокупностью звуков» следует понимать множество звуковых обликов слова в его различных грамматических формах. Следовательно, прежде чем мы установим звучание слова, мы должны знать его грамматические формы. Но, если мы знаем грамматические формы слова, критерий «звучания», в силу своей производности, нам уже не нужен.

Слова «определенное грамматическое употребление» весьма неясны. Мы им придадим, в соответствии с интерпретацией определения А. Мейе в некоторых популярных пособиях, самый определенный смысл, какой только возможен в рамках обычных представлений, и будем считать, что под «грамматическим употреблением» понимается совокупность грамматических форм слова. К сожалению, мы обнаруживаем, что даже понятие грамматической формы слова в традиционной грамматике сформулировано таким образом, что допускает очень большие расхождения во мнениях по поводу того, какие единицы являются формами одного слова. В таком языке, как русский, суффиксы числа и уменьшительности у существительных, аффиксы видообразования и возвратный суффикс *-ся* у глаголов, суффиксы наречий, суффиксы степеней сравнения прилагательных одни ученые считают словоизменяемыми (тогда в парах *дом — дома*, *дом — домик*, *закрывать — закрывать*, *делать — сделать*, *строить — строиться*, *хороший — хорошо*, *хороший — лучше* представлено по одному слову), а другие — словообразовательными (и тогда в тех же парах представлено по два разных слова). Постоянным предметом споров являются так называемые неличные формы глагола: одни ученые включают их в парадигму глагола, а другие считают самостоятельными словами и даже (иногда) самостоятельными частями речи. Понятие грамматической формы является, таким образом, недостаточно простым, чтобы его можно было принять в качестве исходного.

Тогда возникает вопрос о том, с помощью какого критерия мы устанавливаем грамматические формы слова; в частности, вопрос о том, на каком основании мы считаем, что *хороший — лучше* являются формами одного слова. По-видимому, основанием является только то, что эти формы имеют одинаковое лексическое значение; в противном случае мы могли бы с равным успехом объединять в одну парадигму *хороший* и *хуже*, *плохой* и *лучше*. Следовательно, чтобы установить грамматические формы слова, мы должны знать его значение. Понятие грамматической формы является, таким образом, производным от понятия значения.

К сожалению, как показывает практика лингвистических исследований, понятие значения само по себе не является достаточно определенным, чтобы его можно

было использовать в качестве критерия. Не имея строгого описания значений, мы не можем однозначно решать вопрос о том, совместимо ли то или иное семантическое различие в формах с единством слова, или оно нарушает это единство.

Таким образом, исчерпывающее определение слова, учитывающее все его признаки (фонетические, грамматические и семантические), оказалось неудовлетворительным со всех точек зрения. Поэтому многие лингвисты пытаются строить определение слова на основе какого-то одного признака, который выделяется в нем как основной. Чаще других в качестве основного признака слова рассматривались 1) единичность ударения (фонетический признак)¹, 2) потенциальная сводимость слова к предложению (синтаксический признак), 3) неделимость, неразложимость, или цельнооформленность слова (морфологосинтаксический признак). Выгодно отличаясь от определения А. Мейе своей конструктивностью, определения слова, основанные на каком-либо из названных признаков, не соответствуют тем не менее тому множеству объектов, к которым применяется термин «слово».

Действительно, нельзя считать единичность ударения отличительным признаком слова, потому что во многих языках имеются 1) так называемые проклитические и энклитические сочетания слов, произносимые с одним ударением, ср. *Чтобы этого больше не было!* и *Я было думал, Читаем мы много* (см. <2>, <230>); 2) сложные и производные слова, произносимые с двумя ударениями, ср. англ. *unknown* — «неизвестный», рус. *тёмно-коричневый* и т. п. Несостоятельна и попытка соотнести слово с предложением и рассмотреть его как «предельный потенциальный минимум предложения», так как в языке имеется большое количество «слов», никогда не выступающих в качестве предложения, ср. *дескать, мол* и т. п. <30>.

Кроме того, здесь возникает опасность тавтологического круга: приведенная формулировка предполагает, что предложение уже определено независимо от слова; между тем в большинстве существующих определений предложения в той или иной форме используется понятие слова.

¹ Недавно П. С. Кузнецов предложил основанное на фонетическом критерии операционное определение слова <87>.

Даже самый серьезный из этих критериев — морфолого-синтаксический критерий «цельнооформленности», предложенный и исключительно тщательно разработанный А. И. Смирницким <167>, дает основание для конструктивного выделения слова в небольшой группе языков преимущественно флективного строя, таких, как русский или латынь¹. Материалы других языков ясно показывают ограниченный и частный характер учения А. И. Смирницкого.

Ядром этого учения является положение о том, что слово, в отличие от словосочетания, оформляется как целое; сложное слово *овцебык*, например, считается словом, а не словосочетанием на том, в частности, основании, что родительным падежом этого слова является форма *овцебыка*, а не *овцыбыка*; во множественном числе оно имеет форму *овцебыки*, а не *овцыбыки* и т. д. Цельнооформленность слова проявляется и в том, что между частями слова нельзя вставить других слов.

С этой точки зрения отделяемые глагольные приставки в немецком и шведском языках в препозитивном положении ведут себя как часть слова, потому что между приставкой и остальной частью основы не могут помещаться другие слова, а в постпозитивном положении они ведут себя как отдельное слово, поскольку между самостоятельно оформленным глаголом и приставкой могут вклиниваться слова и целые словосочетания, ср. нем. *aufnehmen* — «поднимать» — *nehmen das auf*, швед. *omfatta* — «охватывать» — *fatta det om*. Распространить действие определения и на этот материал можно только ценой отказа от того установившегося взгляда на вещи, что *auf-* и *-nehmen*, *om-* и *-fatta* являются одними и теми же единицами в двух разных случаях употребления, а не парами омонимов (*auf*¹, *om*¹ — приставка; *auf*², *om*² — наречие; *-nehmen*¹, *-fatta*¹ — основа; *nehmen*², *fatta*² — инфинитив). Любое из двух возможных альтернативных решений (либо считать *auf*, *om* и т. д. приставками, либо считать их словами) приводит к противоречиям.

В языках, где невозможна препозиция артикля (армянском, болгарском), он ведет себя с точки зрения кри-

¹ Впрочем, даже в них с известными натяжками; см. глубокий критический разбор взглядов А. И. Смирницкого в работе М. В. Панова <135>.

терия цельнооформленности как часть слова; в языках, где невозможна его постпозиция (английском, немецком, французском, итальянском и др.), он ведет себя как отдельное слово; в языках, где возможна и препозиция и постпозиция артикля, имеет место ситуация, аналогичная только что рассмотренной: в шведском и датском языках постпозитивный, так называемый суффигированный, артикль неотделим от слова, т. е. является его частью, ср. швед. *skogen* — «(этот) лес», а тот же самый артикль, стоящий в препозиции (он употребляется при наличии определяющего существительное с суффигированным артиклем прилагательного), является отдельным словом, ср. *den stora skogen* — «(этот) большой лес». Следовательно, и в этом случае, если мы хотим сохранить верность критерию «цельнооформленности», не впадая при этом в явное противоречие, мы должны расщепить шведский (и датский) определенный артикль на две разные единицы¹.

Необъяснимы, с точки зрения А. И. Смирницкого, и аналитические формы типа рус. *будет работать*, англ. *has written* — «написал», фр. *est venu* — «пришел» и т. п. Они входят в одну парадигму с синтетическими формами типа *работает, writes* — «пишет», *vient* — «приходит» и поэтому считаются «формами слова», но их части оформляются самостоятельно и в предложении разобщаются, поэтому они считаются словосочетаниями. А. И. Смирницкий видел это противоречие, но мирился с ним <169>.

Сознавая неудовлетворительность всех существующих определений слова, многие лингвисты склонялись к мысли, что дать определение этого понятия вообще невозможно. Тем не менее термин, по вполне понятным причинам, пытались сохранить, ссылаясь на то, что любой исследователь интуитивно, но вполне однозначно определяет слово, когда ему приходится иметь с ним дело. К сожалению, конкретная практика лингвистов опровергает это мнение. В лингвистике не прекращаются споры о том, имеем ли мы дело в каком-то конкретном случае со словом, частью слова (морфемой) или словосочетанием. Одни романисты считают французские формы типа *il* — «он», *je* — «я», *le* — «его» и т. п. словами <72>, а другие—

¹ Серьезные трудности возникают в связи с так называемыми «слитными артиклями» или «артикулированными предлогами» ряда романских языков, ср.: фр. *au* (*à+le*), *des* (*de+les*), итал. *al* (*a+il*), *del* (*di+il*), *sul* (*su+il*) и т. п.

приглагольными морфемами <17>, <29>. Одни англисты определяют образования типа *stone wall* — «каменная стена» (буквально «камень стена») как (нестойкое) сложное слово <167>, а другие — как свободное словосочетание <320>. Одни исследователи рассматривают немецкие глаголы с отделяемыми приставками типа *aufnehmen* — «поднимать», *einführen* — «вводить» как слова (такова, в частности, практика немецкой лексикографии), а другие — как фразеологические единицы, т. е. словосочетания <94>. Подчеркнем, что во всех этих случаях речь идет о весьма распространенных формах трех очень хорошо изученных языков.

Сказанное не значит, что мы сомневаемся в полезности понятия слова. Речь идет лишь о том, что рассмотренные нами определения этого понятия не являются точными, поскольку 1) либо в них не указан такой набор допускающих практическую проверку свойств, по которому мы могли бы однозначно относить тот или иной встретившийся нам объект к классу слов или неслов; 2) либо они не соответствуют по объему тому множеству объектов, которые фактически называются данным термином.

В подавляющем большинстве случаев неточность основных лингвистических понятий является прямым и неизбежным следствием того обстоятельства, что они определяются в конечном счете на с е м а н т и ч е с к о й основе. Действительно, поскольку в традиционной лингвистике нет полного формального описания значений, так называемый «семантический критерий» является чисто интуитивным. Конечно, исследователь неизбежно опирается на свое интуитивное знание объекта, когда он приступает к разработке теории. Однако интуиция не может использоваться как решающее средство доказательства, так как в этом случае становятся невозможными отчуждение и однозначная передача знаний: максимум того, что мы можем сделать, построив описание на интуитивной основе, — это апеллировать к интуиции читателя и «возбудить» у него с помощью примеров не в точности совпадающее с нашим, а приблизительно похожее на него представление об объекте.

Задача исследователя, поставившего себе целью создание теории, состоит именно в том, чтобы ф о р м а л и з о в а т ь наши первоначально интуитивные знания об объекте. Только формальная теория допускает экспери-

ментальную проверку и однозначную передачу наших знаний другим лицам.

Семантический критерий ведет себя предательски еще в одном отношении: классификация объектов по их семантическим признакам, не опирающаяся на формальную теорию значения, рискует оказаться бесконечной, потому что на этом пути возможности детализации в принципе безграничны. Сошлемся на весьма показательные в этом отношении рассуждения А. М. Пешковского о значениях падежей в работе «Русский синтаксис в научном освещении»: «Так как значения падежей теснейшим образом связаны с вещественными значениями и управляющих слов и управляемых, то исследователь подвергается соблазну придумать здесь столько рубрик, сколько их можно установить для вещественных значений одного из сочетаемых элементов и другого, прибавив еще рубрики, образуемые комбинациями тех и других случаев. Так, установив, положим, значение *о р у д н о с т и* для творительного падежа в сочетаниях *рубить топором, пилить пилой* и т. д., он может усмотреть новое значение в сочетаниях *схватывать мысль, чувствовать сердцем, понимать умом*, так как здесь и само «орудие» и обращение с ним совершенно иные, и опять-таки новое в сочетаниях *действовать подкупом, добиваться чего силой, терпением, очаровывать кого остроумием* и т. д. В первом случае можно было бы говорить о творительном «умственного орудия», во втором — о творительном «средства»... На этом пути нет предела для дробления значений (например, можно было бы различать «умственное» и «чувственное» орудие, средства физические, экономические, социальные и т. д.)...» (144, 261).

В числе лингвистов, раньше других понявших эти опасности и сделавших наиболее радикальную попытку их преодолеть, был основатель и самый блестящий представитель московской школы лингвистики — акад. Ф. Ф. Фортунатов. Существо взглядов Ф. Ф. Фортунатова по интересующему нас вопросу можно пояснить на примере разработанного им учения о «грамматических классах слов», которое он противопоставил традиционному учению о частях речи.

Традиционный принцип классификации слов по их семантическим, синтаксическим и морфологическим признакам (45) проводится более или менее последовательно

лишь до тех пор, пока семантические, синтаксические и морфологические признаки слова можно считать до известной степени согласованными и предсказывающими друг друга (как это имеет место, например, для существительных, прилагательных и глаголов). Как только эти три признака слова оказываются несогласованными и не предсказывают друг друга, оно ф а к т и ч е с к и классифицируется не по совокупности своих признаков, а только по одному из них — семантическому. Именно на этом основании выделяются в отдельные классы местоимения и числительные, хотя ни тот, ни другой класс не является синтаксически и морфологически однородным. Действительно, слова типа *он, она, никто* ведут себя, в основном, как существительные; слова типа *всякий, каждый, никакой*, относимые к тому же классу, имеют парадигму прилагательных и ведут себя, в основном, как прилагательные. Так называемые количественные числительные, за исключением слова *один*, в русском языке имеют в некоторых условиях синтаксические признаки существительного (ср. *пять девушек, вижу пять девушек*, где они управляют родительным падежом существительного), а в некоторых других условиях — синтаксические признаки прилагательного (ср. *пяти девушкам, пятью девушками*, где они согласуются в числе и падеже с существительным). В противоположность этому, относимые к тому же классу порядковые числительные и слово *один* имеют парадигму и синтаксические функции прилагательных.

Традиционному принципу «многоплановой», а по существу непоследовательно семантической классификации Ф. Ф. Фортунатов противопоставил морфологический принцип, в соответствии с которым все слова делятся на классы по признаку наличия или отсутствия формы (делимости на основу и аффикс) <191, 68>. Таким образом выделяются классы слов I) с формами словоизменения, II) без форм словоизменения. Слова первого класса делятся, в зависимости от типа форм, на (1) слова спрягаемые (глагол), (2) слова склоняемые (существительное) и (3) прилагательные склоняемые слова. Лишь на третьей ступени классификации учитывается семантический признак, на основе которого в классе (2) выделяются три подкласса слов, а в классе (3) — четыре (слова-названия, слова-местоимения, причастия и порядковые числительные). Слова без форм словоизменения (класс II) делятся

на слова с формами словообразования (производные наречия) и слова, не имеющие таких форм (сюда же включаются и непроизводные наречия).

Общие принципы этой замечательной для своего времени классификации оказались настолько жизнеспособными, что почти 50 лет спустя использовались без существенных изменений во многих работах по структурной лингвистике <44>, <251>.

Критика традиционного грамматического учения велась и в других направлениях. Традиционная описательная грамматика является по преимуществу аналитической дисциплиной. Хотя она не разграничивает четко и последовательно точку зрения говорящего и точку зрения слушающего, наиболее распространенный способ описания фактов соответствует последней точке зрения: в качестве исходного материала берутся языковые формы, а результатом лингвистического анализа является перечень возможных значений (или функций) каждой из них. Таким образом построена, в частности, «Грамматика русского языка» (изданная АН СССР <45>), в целом являющаяся хорошим образцом традиционной грамматики; типичным для нее является описание, начинающееся с фиксации формы (грамматической морфемы падежа, числа, времени, склонения и т. п. или типа словосочетания, предложения и т. п.) и заканчивающееся указанием выражаемых ею значений (например, для морфем падежа — значений объекта, инструмента, причины, количества, пространственной и временной протяженности и т. д.). Лишь изредка и несистематически, главным образом при описании словообразовательных суффиксов со значением деятеля, действия, качества и т. п., описание строится в обратном порядке, т. е. от некоторого заданного значения к различным способам его выражения.

Многим крупным ученым, таким, как Г. Пауль <346>, Ф. Брюно <257>, <258>, Э. Сэпир <175>, О. Есперсен <55>, <319>, <320>, Л. В. Щерба <229>, <149>, казалось неоправданным предпочтение, отдаваемое классификационной аналитической грамматике, и они предусматривали в той или иной форме построение грамматик двух типов: грамматики «слушающего» и «говорящего» <55>, грамматики «анализа» и «синтеза» <175>, грамматики «пассивной» и «активной» <229>. Э. Сэпир писал по этому поводу, что к вопросу формы в языке можно под-

ходить с двух различных точек зрения: «Каковы формальные модели языка? и какие типы понятий составляют содержание этих формальных моделей? Эти две точки зрения совершенно различны» <175, 45>.

Ниже у нас будет случай подробно остановиться на вопросе об аналитическом и синтетическом подходе к описанию языка; здесь же, в соответствии с нашим общим планом, мы коротко расскажем о двух наиболее серьезных попытках построить грамматику на новых идейных основаниях — лингвистических учениях Ф. Брюно и О. Есперсена ¹.

Ф. Брюно формулировал свою задачу следующим образом: «Моим желанием было представить методичное изложение фактов мысли, рассмотренных и классифицированных по отношению к языку, и средств выражения, которые им соответствуют» <257, VII>. С этой точки зрения понятен скептицизм Ф. Брюно в отношении господствовавшего в его время типа аналитических грамматик с их склонностью к историзму и увлечением той схемой частей речи, которая сложилась у античных грамматиков на основе фактов греческого и латинского языков. Грамматика Ф. Брюно, идущая от мысли к средствам ее выражения, по самой своей природе исключала и историческую точку зрения на современное состояние языка, и концепцию частей речи: во-первых, исторически меняется не столько мысль, сколько способы ее выражения; во-вторых, одна и та же мысль может быть выражена различными средствами, в том числе различными частями речи. Реформа, намеченная Ф. Брюно, была недостаточно понята и оценена его современниками <268, 12—15>, даже такими выдающимися, как Ш. Балли <241>, который увидел в своем коллеге разрушителя грамматики, в то время как фактически Ф. Брюно достраивал ее.

В свете изложенной выше общей идеи Ф. Брюно группировал, например, в одном разделе своей грамматики «все вопросительные средства и все отрицательные средства, независимо от того, являются ли они прилагательными, местоимениями или наречиями» <258, 252>; все средства выражения причины,

¹ Более полный разбор работ Ф. Брюно и О. Есперсена читатель найдет в обстоятельной статье М. В. Сергиевского <160>, где их взгляды сопоставляются со взглядами В. Брёндаля, Ш. Балли, Л. В. Щербы, Е. Куриловича, В. В. Виноградова.

шающего, О. Есперсен называет морфологией, а второй способ (от значения к форме), соответствующий точке зрения говорящего, — синтаксисом¹. При первом способе в одном разделе грамматики группируются омонимы, при втором — синонимы.

Функции — синтаксические категории — меняются от языка к языку и в каждом языке образуют специфический, свойственный только ему набор, а понятия — внеязыковые, понятийные, или логические, категории — универсальны и независимы от более или менее случайных фактов существующих языков. Поскольку понятийные категории универсальны, каждый язык должен располагать средствами для выражения любой из них, но средства выражения данной категории в разных языках не обязаны совпадать. Более того, поскольку понятийные категории независимы от синтаксических категорий того или иного языка, некоторая понятийная категория может быть выражена различными средствами в пределах одного языка. Так, логическое различие актива и пассива выражается залогом глагола (ср. *to eat* — «есть» — *to be eaten* — «быть съеденным»), суффиксами прилагательных и существительных (ср. *decisive* — «решающий», *talkative* — «разговорчивый» в противоположность *eatable* — «съедобный», *visible* — «видимый», *employer* — «наниматель» в противоположность *employee* — «служащий», «нанятый», *arrival* — «приезд» в противоположность *defeat* — «поражение»; ср. рус. *любитель* — *любимец*, *учитель* — *ученик* и т. п. <320>).

Будучи универсальными, понятийные категории делают возможным построение грамматик на основе общего исходного материала. Грамматический строй любого языка может быть описан как система средств выражения некоторого набора понятийных категорий, неизменного для всех описываемых языков. Такой способ изложения грамматики имеет хотя бы то преимущество, что в данном случае мы располагаем естественной основой для сравнения грамматического строя различных языков и создания научной лингвистической типологии, для которой понятийные категории служат в качестве языка-посредника.

¹ «Синтаксис... смотрит на грамматические факты изнутри... со стороны их значения» <319, 1>.

Учение о понятийных категориях, возвращающее нас к точке зрения так называемых «общих» или «универсальных» грамматик классической эпохи <268, 12—15>, не оказало заметного влияния на лингвистическое мышление современников О. Есперсена¹. Только теперь, когда развернулась практическая работа по составлению синтетических и порождающих грамматик, мы можем вполне оценить значение этих глубоких идей.

Заметный след в истории лингвистической мысли оставила предпринятая рядом языковедов попытка преодолеть э м п и р и з м традиционной описательной грамматики, которая, по словам крупного датского языковеда В. Брендаля, интересуется почти исключительно явлениями, доступными непосредственному наблюдению <27, 40>, и выработать систему достаточно общих понятий, не зависящих от строя того или иного конкретного языка. Такими понятиями могли бы быть понятия об универсальных языковых объектах (обнаруживаемых в строе любого языка) или понятия об объектах, выводимых из особенностей логического строения мысли.

Иллюстрацией поисков общей системы лингвистических понятий может служить известная со времен античности классификация знаменательных слов всего на два класса — имя и глагол. Эта классификация подтверждается и логическими соображениями, и данными конкретных языков. Э. Сэпир считал ее универсальной: «...Какой бы неуловимый характер ни носило в отдельных случаях различие имени и глагола, — писал он, — нет такого языка, который вовсе бы пренебрегал этим различием. Иначе обстоит дело с другими частями речи. Ни одна из них для жизни языка не является повелительной необходимостью» <175, 95>.

Эмпиризм традиционной грамматики проявляется и в том, что она сосредоточивает внимание не на общих, а на с п е ц и ф и ч е с к и х свойствах изучаемых ею объектов. Общие свойства таких, например, единиц, как слово, словосочетание, предложение, обычно не принимаются ею в расчет. С этой точки зрения представляют интерес некоторые идеи О. Есперсена, в частности его учение о синтаксических рангах и типах синтаксической

¹ И. И. Мещанинов <119> пришел к этому учению, по-видимому, независимо от О. Есперсена.

связи, которое служило для него основой единообразной трактовки слов, словосочетаний и предложений.

Все единицы в составе предложения О. Есперсен делил по функциональному (синтаксическому) признаку на единицы трех рангов. В привычных терминах единицы первого ранга — это подлежащие и дополнения; единицы второго ранга — это сказуемые и определения; единицы третьего ранга — различного рода обстоятельства¹; ср.:

We leave here tomorrow — «мы покидаем это место завтра»;
I II I III

a not very cleverly worded idea — «не совсем хорошо сформулированная мысль».
III III III II I

Эту систему О. Есперсен обобщил для сложного предложения, рассматривая главные и придаточные предложения в его составе в качестве единиц первого, второго или третьего ранга <320, 78—90>.

То же стремление к общности и простоте решений характеризует и учение О. Есперсена о двух типах синтаксической связи: юнкции — тесной атрибутивной связи (ср. *красный цветок*) и нексусе — свободной предикативной или полупредикативной связи (ср. *Я заставил его уйти*). Здесь не место излагать детали этого учения; для наших целей достаточно подчеркнуть следующие два момента:

1) В языке есть средства превращения нексусной связи в юнктивную с сохранением ранговой структуры фразы, ср.:

The dog barks furiously — «Пес лает яростно»;
I II III

a furiously barking dog — «яростно лающий пес».
III II I

Таким образом устанавливается эквивалентность (в некотором смысле) словосочетания и предложения.

2) Тот или иной тип связи возможен не только между элементами предложения или словосочетания, но и между элементами производного слова. Так, морфемы слов *cleverness* («умность»), *arrival* («приезд») соединены нексусной

¹ Нельзя не отметить параллелизма между тремя грамматическими рангами и логическими понятиями аргумента (предметной переменной), предиката первого порядка и предиката второго порядка.

связью, в силу чего указанные существительные функционируют как заменители целых предложений¹.

Итак, на основе учения о синтаксических рангах и типах синтаксической связи (нексусе и юнкции) О. Есперсен попытался установить то фундаментальное сходство в строении слов, словосочетаний и предложений, которое позднее послужило одним из краеугольных камней трансформационной теории языка.

* * *

Рассмотренный в этой главе материал показывает, что тенденция к критическому пересмотру лингвистических понятий и приемов исследования языка возникла в н у т р и классического языкознания; в ходе этого пересмотра в ы р а б а т ы в а л и с ь э л е м е н т ы структурного подхода к языку.

Глава 2

НЕПОСРЕДСТВЕННЫЕ ПРЕДШЕСТВЕННИКИ СТРУКТУРНОЙ ЛИНГВИСТИКИ

В главе 1 мы уже касались в определенной мере вопроса о предшественниках структурной лингвистики. К их числу, вне всяких сомнений, принадлежал Ф. Ф. Фортунатов, создавший замечательно стройное для своего времени грамматическое и общелингвистическое учение. Однако по ряду причин ему не суждено было оказать того влияния на развитие науки, на которое мог бы рассчитывать ученый его масштаба <228>. В качестве своих непосредственных предшественников и учителей представители трех основных школ структурной лингвистики называют не Ф. Ф. Фортунатова, а И. А. Бодуэна де Куртенэ и Ф. де Соссюра <27>, <51>, <110>, <179>, <256>, <259>, <376>.

И. А. Бодуэн де Куртенэ², который, по словам Л. В. Щербы, в равной мере принадлежал и русской и

¹ Аналогичные идеи высказывали Л. Теньер <365>, Е. Курилович <89> и некоторые другие исследователи.

² См. о нем <23>, <24>, <25>, <22>, <96>, <223>, <129> и др.

польской науке, во многом предвосхитил открытия Ф. де Соссюра, хотя и не создал цельной концепции, которая могла бы сравниться с соссюровской по законченности, последовательности и ясности. Основные идеи его учения можно суммировать следующим образом:

1. Он обратил внимание на отличие «языка как определенного комплекса известных составных частей и категорий... от языка как беспрерывно повторяющегося процесса» (96, 116). (Эта мысль получила впоследствии четкую формулировку в учении Ф. де Соссюра о языке и речи.) С этим связано предпринятое впервые Бодуэном расщепление понятий звука и фонемы, которые до него не различались. С точки зрения Бодуэна, в языке существует не звук, а фонема, или «звукопредставление», т. е. психическая, а не материальная единица (23, 14—15).

В учении Бодуэна о фонеме достойны внимания следующие четыре пункта:

1) Фонема понималась им, особенно в поздний период его деятельности, как «представление одновременного сложного комплекса произносительных работ» (23, 14), как с у м м а артикуляторных и акустических представлений — кинакем и акусм.

2) Он понимал, что фонемы выполняют различительную функцию. Хотя сами фонемы не имеют значения, они, равно как и «более дробные произносительно-слуховые элементы» (23, 164), «семасиологизуются» и «морфологизуются»; ср. следующие примеры Бодуэна: *сад* — *зад* (семасиологизовано представление работ голосовых связок), *мать* — *дать* (семасиологизована, помимо прочего, работа мягкого нёба).

3) Бодуэн считал нужным разработать такие обозначения, при которых каждому представлению произносительной работы, входящему в состав фонемы, соответствовал бы свой символ.

4) В изучении звуков «Бодуэн стремился к использованию максимально более объективных и точных методов, высказывая при этом гипотезы, поражающие своей прозорливостью (в частности, мысль о таком акустическом исследовании звуков, которое позволило бы представить их визуально)» (70, 142—143).

Именно к этим взглядам Бодуэна восходит современное структурное понимание фонемы как пучка дифференциальных фонологических признаков, метод описания фонем

с помощью так называемых матриц идентификации и экспериментальные спектрографические приемы исследования фонем, позволяющие представить каждую фонему визуально <235>. К сожалению, фонетические и фонологические высказывания Бодуэна, опережавшие состояние современной ему науки, не могли приобрести более конкретного характера из-за отсутствия экспериментальных методов и соответствующей аппаратуры для выделения дифференциальных признаков фонем <235>.

2. В период безраздельного господства сравнительно-исторического языкознания, когда единственным объектом лингвистики, заслуживающим научного внимания, считался исторический процесс, Бодуэн разграничил в языке динамику (процесс) и статику (состояние) и впервые выдвинул мысль о том, что лингвистика должна в равной мере заниматься обеими. «Вне бодуэновской школы,— писал Л. В. Щерба,— все считали, что научность грамматики состоит в ее историчности... теперь, благодаря главным образом Бодуэну, никто не сомневается в том, что описательная грамматика тоже есть предмет науки...» <223, 86—87>.

В связи с этим «стоит и предпочтение, которое Бодуэн всегда отдавал... живым языкам перед мертвыми; на живых языках скорее можно изучить связь явлений, причины их изменений, всю совокупность факторов, управляющих жизнью языка» <223, 89>. При этом в живых языках его интересовала сложившаяся система, а не пережитки некогда существовавших форм и категорий, которые были основным предметом анализа для младограмматиков, так как он понимал, что сложившаяся в данном состоянии языка система может отличаться от системы, характерной для предыдущего состояния языка¹.

3. Бодуэн был одним из первых исследователей, обративших внимание на специфическую структуру письменной речи, отличную от структуры устной речи. Он понимал, что грамматические парадигмы могут иметь различный вид в зависимости от того, какую форму языка — письменную или устную — мы имеем в виду. В связи с этим

¹ К сходным идеям независимо от Бодуэна и приблизительно в одно с ним время пришел Ф. Ф. Фортунатов, отмечавший переход производных основ в непроезводные (ср. *де-ло* → *дел-о*) и различавший понятие корня «для современного языка» (*дел-*) и «в историческом плане» (*де-*) <191>.

Бодуэн разграничил понятие буквы и звука, «благодаря чему многие разделы морфологии приобретают совершенно иной вид, чем тот, к которому мы привыкли в старых грамматиках: *й* в словах *край, май* оказывается не окончанием именительного падежа единственного числа, а составляет неотъемлемую часть основы; то же и в личных прилагательных притяжательных *мой, твой*, которые именительные падежи, оказывается, образуют по именному склонению» <223, 87> (ср. <125>).

4. Бодуэну принадлежат пророческие мысли о будущем лингвистики и о роли, которую займет в ее дальнейшем развитии математика. Как указывает Р. Якобсон, начиная с 70-х годов прошлого столетия Бодуэн обсуждал вопросы непрерывного и дискретного в языке. В 1909 году он выразил убеждение, что лингвистика по мере своего развития будет становиться все ближе к точным наукам. Он предвидел использование в лингвистике количественных и дедуктивных математических методов <360, V>¹.

Учение другого гениального лингвиста, швейцарца Фердинанда де Соссюра, во многом сходно со взглядами Бодуэна. Вместе с тем подчеркнем, что, несмотря на ряд разительных совпадений в концепциях И. А. Бодуэна де Куртенэ и Ф. де Соссюра, честь отчетливой формулировки новых представлений о языке и новых методов исследования языка принадлежит последнему <164>.

1. Величайшая заслуга Ф. де Соссюра состоит в том, что он был одним из первых исследователей, осознавших факт ~~многочисности языка~~. Иными словами, он понял, что ~~язык скрывает не один, а несколько объектов~~. Язык, анализируемый с точки зрения своих функций, может рассматриваться как средство общения, средство выражения мыслей, средство оформления мыслей и т. д. Язык, анализируемый с точки зрения условий своего существования, может рассматриваться

¹ Два десятилетия спустя эти мысли получили вполне деловую формулировку в работах выдающегося ученика Бодуэна Е. Д. Поливанова, который указал на возможность применения дифференциального и интегрального исчисления в экспериментальной фонетике, статистике и теории функций в диалектологии, теории вероятностей в этимологии <145, 177>. Последняя идея была частично реализована в новаторских работах А. Б. Долгопольского <49>.

как факт культурно-исторический. Язык, анализируемый с точки зрения своего внутреннего устройства, может рассматриваться как некоторая знаковая система, служащая для кодирования и декодирования сообщений.

Ф. де Соссюр не только осознал факт многоликости языка, но и выразил это новое представление в ряде созданных им понятий.

2. Ф. де Соссюр противопоставил внешнюю и внутреннюю лингвистику. Внешняя лингвистика изучает условия существования языка, т. е. язык в связи с историей народа и цивилизации, в связи с политикой и литературой, в связи с его географическим распространением и т. д. Внутренняя лингвистика изучает устройство языка, его структуру. Ф. де Соссюр утверждает при этом, что между внутренним устройством языка и внешними условиями его существования нет никакой необходимой или непосредственной связи. Свою мысль он поясняет сравнением языка с шахматами. Тот факт, что шахматы пришли в Европу из Персии, — внешнего порядка. Он никак не определяет системы и правил игры.

3. Внутренний механизм языка можно вполне адекватно изучить и объяснить, ничего не зная о его истории. Более того, плодотворное изучение этого внутреннего механизма предполагает выделение в языке синхронного аспекта, или оси одновременности, в противоположность диахронному аспекту, или оси последовательности. Синхрония связана с диахронией, но не определяется ею. Каждая имеет свой собственный предмет. Синхроническая лингвистика изучает внутреннее устройство языка или его систему, а диахроническая — историю изолированных языковых единиц¹. Здесь, как и в первом случае, помогает сравнение с шахматами. Каждая позиция в ходе игры — это моментальный синхронический срез. «...Любое данное положение характеризуется, между прочим, тем, что совершенно освобождено от всего, что ему предшествовало; совершенно безразлично, каким путем оно установилось; наблюдатель, следивший за всей партией, не имеет ни малейшего преимущества перед тем, кто в критический момент пришел

¹ Мысль Ф. де Соссюра о несистемности диахронии не была принята современной лингвистикой.

взглянуть на состояние игры; для описания данного положения совершенно бесполезно упоминать о том, что произошло десятью секундами раньше» <164, 95>.

Чрезмерное внимание к диахроническим фактам при изучении синхронии может привести к подмене того, что существует в настоящий момент, тем, что имело место в предыдущем состоянии языка. Во французских словах *entier* — «целый», *enfant* — «ребенок» историк выделяет общий префикс *en-*, тождественный латинскому отрицательному префиксу *in-* (ср. *integer* — «нетронутый» и *infans* — «неговорящий»); «субъективный же анализ говорящих полностью его игнорирует» <164, 168>. «Оба вида анализа вполне оправданы, и каждый из них сохраняет свою ценность; но в конечном счете непререкаемое значение имеет только анализ говорящих субъектов, так как он непосредственно базируется на фактах языка» <164, 168>.

С этой мыслью мы уже знакомы в изложении И. А. Бодуэна де Куртенэ и Ф. Ф. Фортунатова (ср. пример со словом *де-л-о* — *дел-о*); ее можно иллюстрировать и следующим ярким примером А. И. Смирницкого. С точки зрения синхронии, т. е. с точки зрения системы отношений языковых единиц в данную эпоху, русское слово *зонтик* есть производное от *зонт*, потому что эта пара входит в словообразовательный ряд *стол* — *столик*, *шар* — *шарик*, *мяч* — *мячик* и т. д. Между тем с диахронической точки зрения, т. е. с точки зрения исторического процесса, *зонт* есть производное от *зонтик*. Это слово, заимствованное из голландского языка (ср. голл. *zonnedeck*), в русском языке было сначала простым по структуре, и лишь позднее оно ассоциировалось с уменьшительными *столик*, *шарик*, *мячик* и др., в результате чего из слова *зонтик* выделилось слово *зонт*.

4. Ф. де Соссюр разделил язык (langue) и речь (parole). Речь связана с языком. Она представляет собой результат использования языка, результат отдельного акта говорения. Существует речь говорящего А, речь говорящего В и пр. Она индивидуальна, линейна, имеет физический характер. Язык — это система взаимосвязанных знаков, обязательная для всех членов данного языкового коллектива. Он социален, нелинеен, имеет психический характер. Язык как система не определяется речью, т. е. индивидуальным использованием этой

системы. Можно было бы развить образ Ф. де Соссюра применительно к этому случаю и сказать, что язык так же независим от речи, как правила шахматной игры — от разыгрывания той или иной партии двумя соперниками. Язык — это правила лингвистической игры, т. е. правила передачи и приема сообщений с помощью некоторой системы знаков. Все носители данного языка обязаны в своей языковой практике подчиняться этим правилам, если они хотят быть участниками эффективного общения.

В свете учения о языке и речи Ф. де Соссюр начал пересматривать понятия о языковых единицах, расщепляя каждое понятие на пару новых понятий: понятие о единице языка и соответствующее ему понятие о единице речи. Так, подобно И. А. Бодуэну де Куртенэ, вместо нерасчлененного термина «звук» Ф. де Соссюр употребляет два четко разграниченных термина: «звук» (единица речи) и «фонема» (единица языка). Эта работа была продолжена последователями Ф. де Соссюра, в том числе и неструктуралистами, и в современной лингвистике часто говорят о фонеме, морфеме, слове, синтагме, типе (или образце) предложения и значении как единицах языка и о звуке (или аллофоне), алломорфе, глоссе (А. И. Смирницкий), словосочетании, предложении и употреблении как соответствующих единицах речи¹.

5. Ф. де Соссюр выдвинул важное положение о с и с т е м н о с т и языка. Язык определяется им как система ~~взаимообусловленных знаков~~. По Ф. де Соссюру, языковой знак состоит из означающего (совокупности фонем, ср. [стол], [teɪbl], [tabl], [tɪf]² и т. п.) и означаемого (понятия «стол»), причем оба элемента знака психического свойства. Ни фонемы, ни значения, взятые по отдельности, не являются языковыми знаками³.

Каждый языковой знак, а следовательно, означающее и означаемое, существуют не сами по себе, а исключительно в силу своего противопоставления другим единицам того же порядка. В языке нет ничего, кроме противо-

¹ См. в особенности <30>, <42>, <52>, <166>, <225>, <246>, <284>, <377>. В русской лингвистической традиции термин «синтагма» используется иначе.

² Транскрипция (безразлично, фонетическая или фонологическая) заключается в квадратные скобки.

³ Похожие идеи высказывали И. А. Бодуэн де Куртенэ <96, 119> и Ф. Ф. Фортунатов <191, 8—10>.

поставлений; «язык есть форма, а не субстанция», — говорит Ф. де Соссюр. Другими словами, существенным для языковой единицы является не материал, из которого она построена, а исключительно множество противопоставлений, в которые она входит. Это множество определяет ее значимость, или ценность. Обратимся в последний раз к соссюровскому примеру с шахматами. Если мы заменим деревянные фигуры фигурами из слоновой кости, металла или стекла, такая замена будет безразлична для системы, потому что отношения между фигурами или их противопоставления друг другу («значимости») остаются прежними. Точно так же если в некотором языке есть противопоставленные друг другу звуки [i], [e], [æ], которые в ходе развития языка изменяются соответственно в [e], [æ], [a] с сохранением прежних противопоставлений, то изменяется лишь физическая субстанция, физический субстрат системы, но не сама система фонем, не лингвистические значимости. Система фонем остается прежней, ибо она не зависит от физического субстрата, в котором реализуется.

Можно представить себе, что система фонем некоторого языка реализуется не в звуковой, а в какой-то другой субстанции, скажем цветовой или тактильной <205>. Каждая фонема данного языка репрезентируется не звуком, а цветом или тактильным ощущением. В таком случае, если мы твердо усвоим систему соответствий тех или иных цветов (или тактильных ощущений) фонемам русского языка, например, мы сможем понять любое сообщение на русском языке, передаваемое потоком цветов или тактильных импульсов ¹.

На первый взгляд, мысль о независимости формы от субстанции и особенно вытекающее из нее положение о том, что человеческий язык может быть не звуковым, кажется фантастической. Однако против нее нельзя привести каких-либо существенных логических соображе-

¹ Это положение не разделяется представителями пражского структурализма <178>, <179>, <236> и некоторыми другими учеными, придерживающимися близких к ним взглядов. По мнению, например, А. Мартине, нормальной первичной субстанцией для языка является звуковая субстанция. Он не отрицает того, что в принципе могут существовать другие субстанции, выполняющие ту же роль в различных коммуникативных системах, но «новая субстанция влечет за собой новую форму (pattern)» <335, 583>.

ний¹, а некоторые экспериментальные данные как будто подтверждают ее.

Вообще говоря, глухой от рождения человек может усвоить систему данного языка, ни разу не слышав, как он звучит. Известно, что глухого сына А. И. Герцена обучали русскому языку на основе зрительных ощущений (по движениям губ). Еще более смелым является эксперимент покойного И. А. Соколянского, который разработал методику обучения языку слепо-глухо-немых. Они усваивали обычный язык (русский) в изобретенном им тактильном коде, а затем переходили на брайлевскую азбуку для слепых. Широко известен подвиг Э. Келлер, слепо-глухо-немой американки, которая таким образом обучилась не одному, а нескольким языкам, занималась историей, литературой и даже математикой.

В подтверждение мысли Ф. де Соссюра о независимости формы от субстанции можно, наконец, сослаться и на обычную практику использования языка. При телефонном разговоре передаваемое нами сообщение последовательно воплощается в следующих различных субстанциях: состояниях нервных клеток мозга, физиологических движениях органов речи, акустических волнах, механических колебаниях мембраны, электромагнитных колебаниях. Во всех этих случаях некоторые структурные характеристики сообщения остаются неизменными, несмотря на то что оно постоянно переходит из одной субстанции в другую. Каждое новое состояние сообщения — образ его предыдущего состояния и в конечном счете образ некоторой внеязыковой ситуации, являющейся предметом сообщения². Из этого примера следует, что строй языка не был бы затронут, если бы тексты на нем записывались не в обычной орфографии, а в механической, электромагнитной или иной «транскрипции».

¹ Мы говорим о логических соображениях, потому что утверждается только логическая независимость формы от субстанции. Это никак не исключает возможности исторической или физиологической зависимости формы от субстанции.

² Ср. следующее глубокое высказывание известного венского логика Л. Витгенштейна: «Граммофонная пластинка, музыкальная мысль, партитура, звуковые волны — все это стоит друг к другу в том же внутреннем образном отношении, какое существует между языком и миром. Все они имеют общую логическую структуру» <38>.

Поскольку языковые единицы являются чисто реляционными сущностями, наличие в языке некоторой единицы можно признать лишь в том случае, если в нем существует противопоставленная единица того же порядка. С этой точки зрения не имело бы смысла говорить о наличии категории транзитивности в языке, в котором есть только транзитивные глаголы и нет нетранзитивных, или о наличии категории падежа в языке, в котором имеется всего один «падеж».

Положение о том, что язык есть форма, а не субстанция, отнюдь не значит, что изучение фонетической или семантической субстанции, в которой реализуется язык, не представляет интереса. Без исследования субстанции нельзя выдвинуть никаких полезных гипотез о лежащей в ее основе системе. Система, не проявляющаяся ни в какой субстанции, не может быть лингвистически интересна. Тем не менее субстанция никак не определяет правил «лингвистической игры»; эти последние логически независимы от физического субстрата, в котором они реализуются.

6. Ф. де Соссюр рассмотрел два вида отношений (противопоставлений) между языковыми единицами: *парадигматические* (Ф. де Соссюр называл их *ассоциативными*) и *синтагматические*. Первые возникают в результате ассоциации единиц по сходству на парадигматической (вертикальной) оси языка, а вторые — в результате ассоциации единиц по смежности на синтагматической (горизонтальной) оси языка, т. е. в речевом потоке.

Ассоциации по сходству возможны на основе значения (ассоциации означаемых, или синонимы, ср. *enseignement* — «преподавание», *éducation* — «воспитание»), на основе звучания (ассоциации означающих, или омонимия, ср. *enseignement* — *justement* — «справедливо») и на основе того и другого (ассоциации означаемых и означающих, ср. *enseignement* — *règlement* — «приведение в порядок», с одной стороны, *enseignement* — *enseignons* — «преподаем» — с другой). Последний случай является наиболее типичным для парадигматических отношений.

Перейдем к синтагматическим отношениям. Примером синтагматических отношений могут служить отношения между морфемами внутри слова, словами внутри словосочетания и т. п. Основным синтагматическим понятием,

которое выдвинул Ф. де Соссюр, было понятие синтагмы. Это понятие, усвоенное без существенных изменений современной структурной лингвистикой, будет подробно обсуждаться в дальнейшем. Здесь достаточно сказать, что понятие синтагмы — гораздо более простое и общее, чем любое сопоставимое с ним понятие традиционной описательной грамматики, — впервые дало лингвисту средство представлять структуру слова и фразы любой степени сложности в виде иерархии единиц одного порядка. Действительно, под синтагмой понимается соединение любых двух элементов (безразлично, морфем, слов или словосочетаний), один из которых является определяемым (главным), а другой — определяющим (второстепенным, зависимым)¹. С этой точки зрения синтагмой является всякое знаменательное слово (за исключением некоторых наречий), поскольку оно состоит минимум из двух морфем, словосочетание, поскольку оно состоит минимум из двух слов, и предложение, поскольку оно сводится либо к слову, либо к словосочетанию. Отметим, что любая синтагма может войти в качестве определяемого или определяющего в состав более сложной синтагмы.

Учением о синтагматике и парадигматике Ф. де Соссюр, по существу, снял традиционное деление грамматики на морфологию и синтаксис и заменил его различием теории парадигм, изучающей отношения языковых единиц на парадигматической оси, и теории синтагм, изучающей отношения лингвистических единиц на синтагматической оси.

7. До сих пор речь шла, в основном, о том, как Ф. де Соссюр понимал язык — объект лингвистики. Суммируем теперь в нескольких словах его представления о лингвистике как науке. Поскольку язык, рассматриваемый с точки зрения своей внутренней организации, представляет собой чисто знаковую, так называемую семиотическую систему, логически независимую от своей манифестации в том или ином субстрате, он сближается с другими знаковыми системами² и вместе с ними составляет

¹ Хотя в этом пункте мы текстуально отступаем от «Курса общей лингвистики», несколько модернизируя его, мы остаемся верными его духу.

² Свойством, отличающим естественный язык от других знаковых систем, является так называемое двойное членение (по терминологии А. Мартине, см. <106>, <335>), состоящее в том, что

объект общей теории знаковых систем, которую Ф. де Соссюр, предвидевший эту науку за несколько десятилетий до ее фактического возникновения, назвал с е м и о т и к о й, или с е м и о л о г и е й. Лингвистика входит в состав семиотики. Как и другие семиотические дисциплины, она является формальной теорией, рассматривающей идеальные объекты, существование которых не выводится непосредственно из наблюдаемых фактов. Поэтому Ф. де Соссюр так охотно сравнивал лингвистику с математикой, которую не интересует физическая природа изучаемых ею объектов. По данным Р. Гodelя, Ф. де Соссюр уже в 1894 году пришел к мысли, что фундаментальные отношения между единицами языка могут регулярно выражаться с помощью математических формул $\langle 281, 44 \rangle$. Впоследствии Ф. де Соссюр сближал лингвистику с алгеброй и геометрией и думал о лингвистических теоремах, которые можно было бы доказывать.

Ф. де Соссюра не раз критиковали за отрыв языка от условий его существования в обществе, синхронии от диахронии, языка от речи, языка от мышления, формы от содержания, парадигматики от синтагматики и т. д. Эта критика не имеет в большинстве случаев достаточных оснований. Как мы видели, Ф. де Соссюр сам указывает на связь языка и общества, синхронии и диахронии, языка и речи, формы и субстанции. Ему принадлежит блестящий образ, иллюстрирующий идею неразрывности языка и мысли: язык и мысль так же неразделимы, как две стороны одного листа бумаги. Он не ограничивается этим и в том же «Курсе общей лингвистики» набрасывает схему «географической лингвистики», «диахронической лингвистики», «лингвистики речи». Однако Ф. де Соссюр действительно пытается вычленить из совокупности разнородных явлений, обнимаемых термином «язык», нечто такое, что выступает как объект собственно лингвистический.

Благодаря предпринятому Ф. де Соссюром расчленению до него диффузно понимавшегося объекта стала возможной столь характерная для других точных и естественных наук интеграция и специализация научных

двусторонние единицы, например морфемы, являющиеся знаками, перекодируются в языке с помощью фонем, т. е. односторонних единиц, которые знаками не являются.

дисциплин. В результате интеграции в точках соприкосновения лингвистики с этнографией и психологией возникли этнолингвистика <238>, <266>, <327>, <331>, психолингвистика <270>, <345> и ряд других гибридных дисциплин. В результате специализации в особую дисциплину выделилась структурная лингвистика, изучающая язык как знаковую систему и не интересующаяся ничем, кроме его кодовых свойств.

Глава 3

КЛАССИЧЕСКИЕ ШКОЛЫ СТРУКТУРНОЙ ЛИНГВИСТИКИ

Известны три классические школы структурной лингвистики, в той или иной мере усвоившие учение Ф. де Соссюра: Пражская во главе с Н. С. Трубецким и Р. Якобсоном (функциональная лингвистика), Копенгагенская во главе с Л. Ельмслевом (глоссематика) и (в меньшей степени) Американская во главе с Л. Блумфильдом и З. Харрисом (дескриптивная или дистрибутивная лингвистика)¹. По отношению к учению Ф. де Соссюра они отличаются друг от друга главным образом акцентами. Так, все школы структурной лингвистики разграничивают синхронию и диахронию и сосредоточиваются на изучении первой, но пражцы более других настаивают на связи синхронии и диахронии и системности последней. Все школы структурной лингвистики разделяют, далее, взгляды Ф. де Соссюра на язык и речь, но при этом пражцы не склонны подчеркивать автономность языка, а американцы и копенгагенцы используют для обозначения этой пары понятий другие

¹ Если считать разновидностями структурализма все лингвистические школы, в той или иной степени порвавшие с традицией, мы должны были бы причислить к ним Женевскую школу (Ш. Балли, А. Сэшэ, А. Фрей), Московскую школу (Р. И. Аванесов, П. С. Кузнецов, А. А. Реформатский, В. Н. Сидоров, А. М. Сухотин) и Лондонскую школу (Дж. Ферс, В. Хаас, М. Холлидей). Однако, хотя каждая из них внесла большой и своеобразный вклад в развитие лингвистики, о них нельзя сказать, без риска впасть в преувеличение, что они разделяли учение Ф. де Соссюра о языке. Поэтому мы не включаем их в наш обзор.

термины («структура» и «текст», «схема» и «текст»). В связи с разграничением языка и речи все школы различают речевые единицы (ср. звуки) и единицы языка (ср. фонемы), причем дисциплины, предметом которых являются речевые единицы (например, фонетика), не считаются лингвистическими. Все школы структурной лингвистики интересуются прежде всего отношениями, оппозициями или функциями элементов в системе, хотя пражцы считают важным и анализ субстанции элементов (фонетической и семантической). Вопреки распространенному мнению, будто дескриптивисты занимаются исключительно синтагматикой, все три школы в той или иной мере исследуют и синтагматические и парадигматические отношения. Наконец, они строят лингвистическое описание в значительной степени на основе общих методологических требований (простоты, полноты, последовательности, объективности, формальности и т. п.). Это и позволяет считать их учения разновидностями структурализма. Более того, хотя области интересов школ, их взгляды на задачи лингвистики, применяемая ими терминология и конкретные приемы исследования языка различны и не допускают синтеза без некоторого насилия над материалом, имеется такая точка зрения, с которой их учения можно рассматривать как дополняющие друг друга (см. стр. 77).

Ниже дается очерк основных положений американской, копенгагенской и пражской школ структурной лингвистики ¹.

Два классика — Э. Сэпир и Л. Блумфильд — стоят у истоков американской лингвистической школы ². От их учений берут начало две ветви американской лингвистики, из которых лишь одна, восходящая, в основном, к Л. Блумфильду, представляет собой разновидность соссюрианского структурализма. Другая, восходящая к Э. Сэпиру, ветвь выходит за его пределы, поскольку в учениях последователей Э. Сэпира

¹ Подробности читатель найдет в работах <27>, <51>, <65>, <66>, <95>, <104>, <105>, <156>, <110>, <162>, <172>, <178>, <44>, <129>, <176>, <196>, <238>, <251>, <269>, <295>, <304>, <356>, <366>, <371>. В большинстве названных работ имеются ценные библиографические указания.

² Непосредственным предшественником обоих языковедов, оказавшим на них большое влияние, был Ф. Боас — первый американский лингвист, предпринявший научное описание индейских языков Америки <254>.

результаты структурного анализа языка сопоставляются с результатами структурного анализа всей материальной и духовной культуры народа, являющегося носителем этого языка. Тем не менее собственно лингвистические идеи Э. Сэпира представляют и теперь значительный интерес и поэтому будут кратко рассмотрены ниже.

Исходные положения учения Э. Сэпира очень близки концепциям Ф. де Соссюра, с которыми мы познакомились в предыдущей главе. Подобно Ф. де Соссюру, Э. Сэпир различает в языке физическую и идеальную систему (модель), причем именно последнюю он считает «реальным и наисущественнейшим началом в жизни языка» (175, 44). Так, фонологическая модель, определяющая число, соотношение и функционирование фонетических элементов, может сохраняться на долгое время и после изменения своего фонетического содержания. «Может случиться, что у двух исторически родственных языков или диалектов нет ни одного общего звука, а модели их идеальных звуковых систем могут оказаться тождественными» (175, 44). С другой стороны, два языка могут иметь идентичные звуки, но совершенно разные фонологические модели. Это касается не только фонологического, но и всех других уровней языка, включая семантический; у каждого языка свой «покрой», и хотя исходная материя может быть одной и той же для многих языков (как один и тот же материал для многих костюмов), в различных языках она раскраивается различным образом¹.

С этими идеями связана широко известная «гипотеза лингвистической относительности» Э. Сэпира и Б. Уорфа (182), (183), (184), не имеющая параллелей в учении Ф. де Соссюра. Два ее важнейших положения получили четкую формулировку в более поздних изложениях, особенно в работах выдающегося американского лингвиста и антрополога Г. Хойера (312), (313), (314):

1) Поскольку у каждого языка есть некая неповторимая «модель», по которой он скроен, каждый язык по-свое-

¹ С этим связано введенное позднее М. Сводешем (363) и К. Пайком (349) деление каждой лингвистической дисциплины на две новые: «этическую дисциплину», изучающую фонетический (или семантический) субстрат модели, и «эмическую дисциплину», изучающую самую модель. Термины «этический» и «эмический» образованы от суффиксов слов «фонетика» и «фонемика» (американское обозначение фонологии). В соответствии с этим различаются этические единицы (звуки) и эмические единицы (фонемы).

му членит действительность и навязывает этот способ членения мира всем говорящим на нем людям. Язык оформляет мысль: люди, говорящие на разных языках, видят мир по-разному. Грамматический строй языка нутка вынуждает говорящего каждый раз, когда он упоминает кого-либо или обращается к кому-либо, указывать, является ли это лицо левшой, лысым, низкорослым, обладает ли оно астигматизмом и большим аппетитом <326, 900—901>. Язык нутка заставляет говорящего мыслить все эти свойства совершенно независимо от того, считает ли он соответствующую информацию существенной для своего сообщения или нет.

2) Языковые модели связаны с культурно-социальными¹. Грамматические и лексические различия, обязательные в данном языке, соответствуют различиям поведения, обязательным в данной культуре. Так, глагольная система языка навахо резко отличается от обычной в европейских языках системы обилием категорий, описывающих все аспекты движения и действия. Грамматические категории навахского глагола заставляют классифицировать в качестве разных объектов движение одного тела, двух тел, более чем двух тел, а также движение тел, различных по форме и распределению в пространстве. Даже предметные понятия выражаются не прямо, но через глагольную основу, и поэтому предметы мыслятся не как таковые, но как связанные с определенным видом движения или действия. По мнению Г. Хойера (на материалы которого мы здесь опираемся), господствующей концепцией, пронизывающей весь строй языка навахо, является концепция мира, находящегося в движении. «Параллели к этой семантической теме,— пишет Г. Хойер,— могут быть найдены в области культуры навахо, рассматриваемой в целом. Даже в наши дни навахо являются по преимуществу бродячим, кочевым народом, перегоняющим свой скот с одного пастбища на другое. Мифы и легенды очень четко отражают этот мотив: боги и герои сказаний без устали путешествуют с одного святого места на другое» <312>. Подчеркнем, что когда говорят о связи языка и культуры в смысле Б. Л. Уорфа, то имеют в виду соответствие строя языка строю культуры в целом. С этой точки зрения не имеет никакой доказа-

¹ Этого тезиса Э. Сэпир не разделял <175, 171—172>.

тельной силы тот давно известный и довольно поверхностный факт, что, например, в языках народов Севера чрезвычайно много наименований для разновидностей снега (при отсутствии родового термина «снег»), у горцев — наименований для различных видов гор и т. п.

Гипотеза лингвистической относительности не стала теорией, потому что до сих пор не удалось найти убедительных экспериментов, которые подтверждали бы ее основные положения. Однако она стимулировала чрезвычайно интересные экспериментальные исследования в области, ранее совершенно не разрабатывавшейся ни лингвистами, ни психологами.

В учении Ф. де Соссюра нет параллелей и для еще одной интересной темы учения Э. Сэпира, которая представляет собой более положительный вклад в науку, чем гипотеза лингвистической относительности. Своеобразие формы («покроя») в языке он считал вторичным, хотя и весьма характерным для языка явлением; в принципе «формальные противопоставления» должны соответствовать «концептуальным различиям», которые Э. Сэпир, в отличие от Б. Уорфа, склонен был, по-видимому, считать универсальными. Недаром он сравнивал переход от языка к языку с переходом от одного геометрического способа представления данной, одной и той же вещи к другому. Однако в естественных языках идеал нарушается. Разбирая латинскую фразу *illa alba femina quae venit* («та белая женщина, которая приходит»), Э. Сэпир указывает, что логически только падеж требует в ней выражения; остальные грамматические категории (род, число в указательных и относительных словах, в прилагательном и глаголе) совершенно не нужны или (число в имени, лицо, время) не относятся к существу синтаксической формы предложения. Пользуясь позднее введенными понятиями, мы могли бы сказать, что логически язык тем совершеннее, чем меньше доля выражаемой в высказывании обязательной информации, вынуждаемой исключительно правилами кодирования, а не существом сообщаемого. С этой точки зрения латинский язык далек от логического совершенства. Действительно, если учесть всю информацию, выраженную в приведенной выше фразе, мы получим нечто вроде «тот — ж е н с к и й — о д и н — д е я т е л ь б е л ы й — ж е н с к и й — о д и н — д е я т е л ь ж е н щ и н а — о д н а — д е я т е л ь,

который — женский — один — деятель — приходит — один — сейчас — действительно...». Мы выделили части высказывания, несущие избыточную, несущественную для сообщаемого информацию, выражение которой оказывается вынужденным обязательными в латинском языке правилами кодирования сообщений, т. е. грамматическими правилами (ср. <113>).

Формальные противопоставления, не соответствующие концептуальным различиям, но требующие обязательного выражения при построении высказывания и поэтому придающие своеобразие грамматическому строю языка, играют роль шумов, в определенной мере искажающих логическую основу высказывания, но не разрушающих ее совершенно. С этой точки зрения представляет большой интерес предпринятая Э. Сэпиром классификация выражаемых в языке понятий на 1) конкретные (ср. *дом-, бел-, бег-*), 2) деривационные (ср. *учи-те-л-я, собра-ни-е*), 3) конкретно-реляционные (ср. род и число в указательных и относительных словах, прилагательных и глаголах) и 4) чисто-реляционные понятия (ср. падеж существительного). Первые и последние, говорит Э. Сэпир, должны быть выражены обязательно; без этого невозможно построить высказывание ни в одном языке: не существует языков без лексики и синтаксиса, хотя существуют языки без морфологии. Языки, в которых выражаются только эти два типа понятий, называются чисто-реляционными. Они «ближе всего подходят к самой сути языкового выражения» <175>.

Это, может быть не всегда осознанное, стремление добраться до логической основы высказывания, вскрыть в изменчивых фактах грамматик универсальный и надежный логический фундамент, найти такие исходные лингвистические понятия, которые поддаются логическому обоснованию, можно проследить и в других положениях учения Э. Сэпира, в частности в уже известной нам критике традиционной классификации слов по частям речи, которую он хотел заменить наблюдаемым во всех языках противопоставлением имени и глагола.

Л. Блумфильд, другой классик американской лингвистики, был в некотором смысле прямой противоположностью Э. Сэпира. Э. Сэпир обладал поразительной лингвистической интуицией, которой мы обязаны его блестящими гипотезами. Однако он относился весьма беззаботно к

форме изложения своих идей; язык сэпировской лингвистики удивительно неточен. В противоположность этому гораздо более скромные гипотезы и идеи Л. Блумфильда изложены на языке, который в те годы (более 30 лет назад) был в лингвистике образцом научной точности <252>, <253>. Л. Блумфильд поставил себе целью разработать единую, конструктивную и последовательно построенную систему понятий, годных для синхронного описания языка любого строя, и хотя сейчас многие детали этой системы представляют лишь исторический интерес, она до сих пор не утратила своего значения. Мы не будем излагать ее целиком, но сосредоточим внимание лишь на тех ее принципах, которые существенным образом повлияли на дальнейшее развитие лингвистики.

1. Подобно Ф. де Соссюру, Л. Блумфильд попытался вычлени́ть из того сложного комплекса явлений, который обозначается термином «язык», объект собственно лингвистический. С его точки зрения, объектом лингвистики являются не звуки и значения сами по себе, а «сочетание определенных звуков с определенным значением» <253, 27>. Звуки интересуют лингвиста лишь постольку, поскольку они различают значения; существенными признаками звуков (фонем) языка являются те признаки, с которыми связывается различие в значении.

Точно так же в грамматике и лексикологии лингвиста должно интересовать не конкретное значение формы и слова, которое, по мнению Л. Блумфильда, «не может быть проанализировано в рамках нашей науки» <253, 161>, а факт различия значений двух форм или слов. Таким образом, Л. Блумфильд ограничил роль «семантического критерия» в лингвистике использованием дифференциального значения; этот принцип, вполне эквивалентный принципу «коммутации» Л. Ельмслева (см. стр. 56), был позднее усвоен многими американскими дескриптивистами.

2. Формы, в которых определенные звуки сочетаются с определенным значением, Л. Блумфильд считал языковыми (в отличие от фонем, которые языковыми формами не являются). Все языковые формы он делил, во-первых, на связанные, никогда не произносимые отдельно (морфемы и другие части слова), и свободные, произносимые отдельно от других форм (слова, сочетания слов и т. п.); во-вторых, на сложные, имеющие

частичное фонетико-семантическое сходство с другими языковыми формами (ср. слово, словосочетание, предложение), и простые, не имеющие такого сходства (морфемы). Из этих двух классификаций Л. Блумфильд извлекает определения всех выработанных им лингвистических понятий, в частности важнейших понятий составляющей, класса и конструкции.

3. Общая часть любых двух сложных форм, являющаяся языковой формой, представляет собой конституент, или компонент, этих сложных форм. Конституенты делятся на непосредственно составляющие и конечные (терминальные) составляющие, которыми являются морфемы. Понятие непосредственно составляющих, близкое сосюровскому понятию синтагмы, иллюстрируется следующим примером: *Poor John ran away* («Бедный Джон убежал прочь»); это предложение делится на две непосредственно составляющих — *poor John* и *ran away*, — каждая из которых в свою очередь делится на две новых непосредственно составляющих (*poor* и *John*, *ran* и *away*) и т. д., пока мы не дойдем до отдельных морфем. Понятие непосредственно составляющих на долгие годы определило направление формальных синтаксических исследований и было с успехом использовано почти 30 лет спустя во многих машинных грамматиках и математических моделях языка (см. части III—V).

4. Языковая форма, которая в определенных условиях заменяет любую форму из некоторого множества форм, называется субститутом. Субституты образуют класс форм.

5. Языковые формы, в которых ни одно из непосредственно составляющих не является связанной формой, называются синтаксическими конструкциями. Существует два основных типа синтаксических конструкций: экзоцентрические и эндоцентрические. Если фраза принадлежит к тому же классу форм, что и какая-либо из ее составляющих, она является эндоцентрической (ср. *poor John*, заменяемое на *John* и потому относимое к тому же классу форм). В противном случае мы имеем дело с экзоцентрической конструкцией (ср. *John ran*).

Такова в основных чертах разработанная Л. Блумфильдом система определений лингвистических понятий.

Многие из них, в частности понятия непосредственно составляющих, субституции, экзоцентрической и эндоцентрической конструкции, получили признание и за пределами структурной лингвистики. Система в целом стимулировала колоссальную методологическую работу по совершенствованию техники лингвистического анализа и явилась фундаментом, на котором строилась с конца 30-х и до начала 50-х годов так называемая дистрибутивная лингвистика — наиболее широко известная и авторитетная разновидность американского структурализма.

Наиболее крупными исследователями, разрабатывавшими в той или иной мере идейное наследие Л. Блумфильда, являются Б. Блок, Е. Найда, Дж. Трейджер, З. Харрис, Ч. Хоккет и некоторые другие. Каждый из них, в особенности З. Харрис, Е. Найда и Ч. Хоккет, внес большой и оригинальный вклад в развитие американского структурализма, но рассмотреть их концепции по отдельности мы, за недостатком места, не сумеем. Ниже мы ограничимся обзором основных положений дистрибутивной лингвистики, так или иначе разделяемых или разделявшихся большинством названных выше ученых.

Американская дистрибутивная лингвистика не является теорией языка в обычном смысле этого слова. По свидетельству Р. Уэллса, она представляет собой «набор предписаний об описании» (378, 38). В качестве такового она может рассматриваться как схема процессов, ведущих к открытию грамматики некоторого языка, или как экспериментальная техника сбора и первоначальной обработки сырых данных.

Основные экспериментальные приемы дистрибутивной лингвистики возникли из того продиктованного изучением индейских языков убеждения, что лингвист подобен дешифровщику¹ или естествоиспытателю, не имеющему никакой заранее заданной информации об объекте, который он собирается изучать. Единственной реальностью, с которой лингвист имеет дело, является текст, подлежащий «дешифровке». Все сведения о «коде» (языке), лежа-

¹ «Дешифровочный» подход к языку возник у американских дескриптивистов и под влиянием некоторых обстоятельств чисто личного свойства, в частности того факта, что в период второй мировой войны младшее поколение американских лингвистов участвовало в работе по шифровке и дешифровке военных документов.

щем в основе этого текста, должны быть выведены исключительно из анализа последнего. Но в тексте непосредственно не содержатся данные о значениях слов языка, его грамматике, его истории и генетических связях с другими языками; непосредственно в тексте даны лишь некоторые его элементы (части, отрезки), и для каждого из них мы можем установить распределение или дистрибуцию — «сумму всех окружений, в которых он встречается, т. е. сумму всех (различных) позиций элемента относительно других элементов» (295, 15—16). Поэтому анализ дистрибуции элементов, и только он, дает нам возможность извлечь из текста искомые сведения о языке.

С этим связаны следующие особенности американской дескриптивной лингвистики как разновидности структурализма:

1) Представление о лингвистическом описании как наборе независимых от строя того или иного конкретного языка процедур обработки текстов, выполнение которых в определенном порядке должно автоматически привести к открытию грамматики (структуры) данного языка. Структура каждого языка принципиально неповторима (это, между прочим, является дополнительным аргументом в пользу того, чтобы изучать ее с чисто внутренней точки зрения, не искажая действительного положения вещей под влиянием внешней информации о языке); однако обнаруживается она, как мы бы сказали сейчас, в результате применения к тексту единой серии универсальных алгоритмов.

2) Различение в языке нескольких уровней: фонологического, морфологического, а в последнее время еще и синтаксического. Эти уровни образуют иерархию, основанием которой является фонологический уровень, а вершиной — синтаксический. Единицы более высокого уровня строятся из единиц непосредственно предшествующего уровня (морфемы — из последовательностей фонем, конструкции — из последовательностей морфем или символов классов морфем). Поэтому, приступая к описанию языка, мы должны начать с обнаружения его простейших единиц и переходить ко все более сложным единицам. Если это правило не будет соблюдено и мы, не найдя всех единиц низшего уровня, перейдем на высший, то либо наше описание не будет полным, либо в определениях воз-

никнет тавтологический круг: морфемы будут определены через фонемы, а фонемы, по крайней мере частично, — через морфемы¹.

3) Представление о единицах языка как классах в некотором смысле дистрибутивно эквивалентных единиц текста (вариантов данной языковой единицы). Этот принцип отличает дистрибутивную лингвистику от школ Ф. де Соссюра и Э. Сэпира, которые считали, что единицы языка принципиально невыводимы из текстовых данных (ср. <175>, <164>).

4) Требование объективности описания, которое при дешифровочном подходе к языку является залогом и единственной гарантией истинности наших знаний. Это требование отличает дескриптивистов от представителей копенгагенской школы структурализма, которые стремятся не столько к объективности описания (независимости его результатов от исследователя), сколько к его формальности (однозначности).

Описать структуру языка исчерпывающим образом — значит установить 1) его элементарные единицы на всех уровнях анализа, 2) классы элементарных единиц, 3) законы сочетания элементов различных классов. Элементарными единицами языка считаются фонемы на фонологическом уровне и морфемы на морфологическом, изучаемые соответственно фонологией (фонемикой) и морфологией. Внутри фонологии выделяется фонотактика — наука о законах связи фонем, занимающаяся исследованием ограничений, налагаемых данной структурой на возможности их сочетаемости. Внутри морфологии выделяется аналогичная морфологическая дисциплина — морфотактика, которая, по существу, совпадает с синтаксисом. С observance этой дисциплины был связан первоначальный отказ дескриптивистов от обычного деления грамматики на морфологию и синтаксис. На стыке между фонологией и морфологией располагается морфонология (морфофонология), предметом которой является изучение фонологических способов выражения морфем (в нее входит и все учение о регулярных фонологи-

¹ См. критику этого принципа у К. Пайка <347>, <348>; ср. взгляды П. С. Кузнецова <84>, М. Холлидея <129>, <287>, Е. Бенвениста <247>, Н. Хомского <199>.

ческих и морфологических чередованиях). Она либо включается в фонологию или морфологию, либо выделяется в самостоятельную лингвистическую дисциплину. Конкретные (лексические) значения тех или иных единиц, а также их звуковой состав не входят в структуру языка, и поэтому изучение семантики и фонетики в узком смысле не входит в задачи лингвистики.

Решение первой из названных выше задач, т. е. выделение элементарных единиц языка, достигается с помощью экспериментальной техники сегментации текста и дистрибутивного анализа текстовых единиц, обнаруженных в результате сегментации. Классы элементарных единиц строятся на основе экспериментальной техники сублиституции (замещения), а законы сочетания элементов различных классов устанавливаются с помощью анализа по непосредственно составляющим. Первые три техники могут применяться, по мнению их создателей, для анализа любого аспекта любого языка, а анализ по непосредственно составляющим имеет силу только в области морфологии и синтаксиса.

Идея о том, что единицы языка, классы единиц и связи между единицами могут быть определены исключительно через их окружение, т. е., говоря словами Ф. де Соссюра, через их отношения к другим единицам того же порядка, и составляет существо дистрибутивного подхода к языку¹.

К сожалению, принципы дистрибутивного анализа не были реализованы американскими дескриптивистами с достаточной последовательностью. В своей практической работе большинство дистрибутивистов отступают от провозглашенной в качестве идеала стратегии лингвистического исследования. Это проявляется прежде всего в том, что фактически лингвист имеет дело не с текстом, а с информантом, т. е. лицом, для которого изучаемый язык является родным (вообще говоря, таким лицом может быть и сам исследователь). Информант должен давать ответы «да» или «нет» на следующие два типа вопросов: (1) правильна ли предъявляемая форма (говорят так или нет)? (2) Являются ли две предъявляемые формы тождественными по значению или различными?

¹ На роль окружения (понимаемого, правда, несколько шире, чем у дистрибутивистов) для отождествления и разграничения языковых единиц указывал еще Ш. Балли <242>.

Обращение к информанту было бы совершенно законным техническим приемом, если бы ему задавались только такие вопросы, ответ на которые можно было бы получить и из текста, но ценой нерационального увеличения объема исследования. К их числу можно отнести вопросы типа (1), так как понятие правильной формы отождествимо, по крайней мере в задачах, подобных обсуждаемой, с понятием «формы, встречающейся в текстах с определенной, достаточно большой вероятностью». В этом случае обращение к информанту действительно сокращает решение задачи, ничего не меняя в ней по существу. Вопросы типа (2) совсем иного рода. Текст не содержит в явном виде сведений о различиях и сходстве значений тех или иных форм, и поэтому, требуя от информанта ответа на вопросы второго типа, исследователь-дистрибутивист подменяет первоначально поставленную, трудную, но исключительно интересную задачу другой задачей, менее трудной и интересной.

С помощью информанта, сообщающего необходимые лингвисту сведения о «дифференциальных значениях», решается первая и самая важная из названных выше задач — задача выделения элементарных единиц языка на всех уровнях анализа. Для решения этой задачи необходимо, как мы помним, прежде всего сегментировать текст, т. е. разбить его на отрезки, являющиеся элементарными единицами того или иного уровня. Посмотрим, например, как производится сегментация текста на морфологическом уровне. Элементарными текстовыми единицами этого уровня являются м о р ф ы, которые в работах З. Харриса, Ч. Хоккета и Е. Найда, написанных в 40-х годах <288>, <305>, <342>, <343>, определяются, с незначительными вариациями в формулировках, как мельчайшие последовательности фонем, регулярно встречающиеся в различных участках текста с одним и тем же различием в значении (Е. Найда) или с одним и тем же значением (З. Харрис, Ч. Хоккет). Это, по существу, мало чем отличается от обычных семантических определений. В других работах содержатся замечания о том, что морфы можно выделить без обращения к значению — методом последовательных приближений (методом «проб и ошибок»), но строго этот метод дескриптивистами нигде не описан.

Нельзя признать удачной и несемантическую технику выделения морфов, предложенную З. Харрисом в 1951 го-

ду <295, 158—160>. З. Харрис рассматривает последовательности фонем *A, B, C* и *D* и некоторое общее окружение *X*. Необходимым условием для признания *A, B, C* и *D* морфами является наличие комбинаций хотя бы следующих трех типов: *ABX, ADX* и *CDX*. Тогда в паре *ABX* и *ADX* можно выделить морфы *B* и *D*, а в паре *ADX* и *CDX* — морфы *A* и *C*. Однако это условие не является достаточным, так как не запрещает нам выделить в качестве морфа отдельную фонему, ср. *Where's the — (ba-g, ru-g, bu-g)?* аналогичный русский пример, *где — (ма-к, лу-к, со-к)?* Поэтому выдвигается второе условие, которое, к сожалению, нельзя считать сколько-нибудь определенным: последовательности фонем рассматриваются как морфы только в том случае, если «многие из этих последовательностей имеют идентичные отношения ко многим другим предположительно независимым последовательностям фонем» <295, 160>¹.

Сегментирование текста на элементарные единицы (звуки или фонемы на фонологическом уровне и морфы на морфологическом) является лишь первым шагом в процедуре выделения соответствующих единиц языка — фонем или морфем. Вторым шагом является и д е н т и ф и к а ц и я — установление того, какие из элементарных текстовых единиц тождественны между собой, т. е. являются вариантами одной и той же единицы языка (аллофонами одной фонемы или алломорфами одной морфемы), а какие различны, т. е. являются представителями разных единиц языка. Для решения этой задачи используется дистрибутивный анализ в собственном смысле слова.

Устанавливаются три типа дистрибуции элементов:

1) Текстовые единицы находятся в д о п о л н и т е л ь н о й д и с т р и б у ц и и, если они никогда не встречаются в одинаковых окружениях. Этого условия в большинстве случаев достаточно, чтобы признать ряд звуков вариантами (аллофонами) одной фонемы. Таковы различающиеся степенью закрытости — открытости гласные звуки в словах *семь, день* (наиболее закрытый вариант, произносимый в позиции после мягкой согласной и перед мягкой согласной); *сел, дел* (более открытый вариант,

¹ Позднее З. Харрис разработал гораздо более серьезную (хотя и не лишенную недостатков) несемантическую технику выделения морфов. Она обсуждается нами в части III книги.

произносимый после мягкой и перед твердой); *шесть*, *жесть* (еще более открытый вариант, произносимый после твердой и перед мягкой); *шесть*, *жест* (наиболее открытый вариант, произносимый после твердой и перед твердой). Для признания ряда морфов алломорфами одной морфемы требуется, чтобы они не только были в отношении дополнительной дистрибуции, но и имели одно и то же дифференциальное значение. С этой точки зрения алломорфами одной морфемы должны быть признаны вспомогательные глаголы *être* — «быть» и *avoir* — «иметь» во французском языке, *sein* — «быть» и *haben* — «иметь» в немецком в тех (и только тех) случаях, когда они служат для образования аналитических форм времени (*passé composé* и других во французском языке, *Perfekt*'а и других в немецком): один класс глаголов образует это время с помощью *être* (*sein*), а другой, непересекающийся с ним класс — с помощью *avoir* (*haben*)¹.

2) Текстовые единицы находятся в контрастной дистрибуции, если они могут встречаться в одних и тех же окружениях, различая значения. В этом случае они являются представителями разных единиц (фонем или морфем). Таковы начальные согласные в словах *том* — *дом* — *ком* — *лом* — *ром*, принадлежащие пяти различным фонемам. В контрастной дистрибуции находятся морфы *-ed* и *-ing* в английском языке, принадлежащие, следовательно, различным морфемам (ср. *She was charmed* — «Она была очарована» и *She was charming* — «Она была очаровательна»). Заметим, что для установления фонологического различия достаточно одной, так называемой минимальной, пары с контрастом; ср. единственную в английском языке пару *Aleutian* — «алеутский» — *allusion* — «намек», в которой выделяются фонемы [ʃ] и [ʒ] (пример из книги Г. Глисона <44>).

3) Текстовые единицы находятся в свободном чередовании, если они встречаются в одних и тех же окружениях, не различая значений. В этом случае они являются вариантами одной и той же единицы языка. Примером свободного чередования является переднеязычное (вибрирующее) и увулярное (грассированное) [ɾ]

¹ Глаголы, которые в одних лексических значениях образуют эти времена с помощью *être* (*sein*), а в других — с помощью *avoir* (*haben*), должны рассматриваться с этой точки зрения как пары омонимов.

во французском языке (вибрирующее [r] является нормой сценического произношения). В свободном чередовании находятся варианты морфемы творительного падежа *-ей* и *-ею* в русском языке, ср. *землей* и *землею*.

Мы видим, таким образом, что при идентификации единиц языка, точно так же, как при сегментировании текста, имеет место отступление от чисто дистрибутивных процедур и использование по существу семантического критерия дифференциального значения. Упомянем в связи с этим попытку Б. Блока <249> поправить положение дел введением строгих определений трех типов дистрибуции на основе некоторых элементарных понятий математической теории множеств¹. По мнению Б. Блока, существует четыре типа распределения единиц текста относительно друг друга: 1) две текстовые единицы никогда не встречаются в одинаковых окружениях; 2) две текстовые единицы всегда встречаются в одних и тех же окружениях; 3) множество окружений, в которых встречается одна из единиц, целиком входит в множество окружений, в которых встречается вторая; 4) множества окружений, в которых встречаются текстовые единицы, частично пересекаются. Если обозначить кружком множество окружений, в которых встречается некоторая единица текста, то эти четыре типа распределения можно иллюстрировать следующим рисунком:

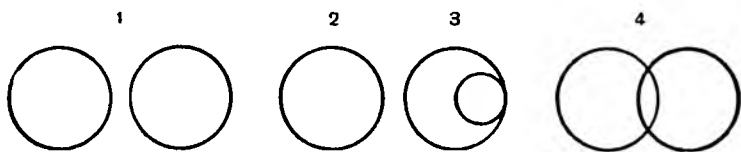


Рис. 1.

Первый случай распределения Б. Блок предлагает называть дополнительной дистрибуцией, второй — свободным чередованием, третий и четвертый — контрастной дистрибуцией.

К сожалению, это уточнение не устраняет коренного

¹ Используемые Б. Блоком понятия теории множеств понадобятся нам в дальнейшем; поэтому их формальные определения излагаются ниже, во II части, а здесь мы ограничиваемся содержательным комментарием.

недостатка классического дистрибутивного анализа. Рассмотрим на фонологическом примере две ситуации, делающие невозможным применение предложенных Б. Блоком критериев:

1) В английском языке глухие взрывные непридыхательные звуки [p], [t], [k] находятся в дополнительном распределении сразу с двумя рядами звуков: глухими взрывными придыхательными [p^h], [t^h], [k^h] и звонкими взрывными [b], [d], [g]. Глухие взрывные придыхательные и звонкие взрывные относятся к разным фонемам, так как они контрастируют в ряде окружений. Звуки [p], [t], [k] могут, следовательно, считаться либо вариантами фонем [p], [t], [k], либо вариантами фонем [b], [d], [g] соответственно. Поскольку критерий дополнительной дистрибуции не дает нам оснований для решения этого вопроса, в равной мере допуская обе возможности, мы вынуждены обратиться к исключенному ранее недистрибутивному критерию «фонетического сходства» <295>, с помощью которого мы объединяем [p] с [p^h], [t] с [t^h] и [k] с [k^h] в качестве вариантов одной фонемы¹.

2) В английском, немецком, датском и шведском языках имеются звуки, находящиеся в дополнительной дистрибуции и тем не менее не считающиеся вариантами одной и той же фонемы. Таковы [h] и [ŋ]. Для того чтобы запретить объединение [h] с [ŋ] по правилу дополнительной дистрибуции, выдвигаются новые частные критерии.

В результате ряда отступлений от чисто дистрибутивных процедур анализ лишается объективности, и на одном и том же материале разные выводы получают не только разные ученые, но и одни и те же исследователи в разные периоды своей деятельности. Э. Хауген указывает, например, на следующие любопытные факты: Дж. Трейджер при анализе американского варианта английского языка в 1940 году выделил шесть гласных фонем, а в 1947 — девять. М. Сводеш в 1935 году рассматривал американские варианты английских дифтонгов как единичные фо-

¹ Даже здесь возникают технические трудности. По авторитетному свидетельству Д. Джоунза <321>, непридыхательное [p], если оно произносится в словах типа [pɑ:k], на слух англичанина ближе к [b], чем к [p], и может быть принято им за звонкий (т. е. он может услышать [bɑ:k] вместо [pɑ:k]). Поэтому, помимо критериев фонетического сходства, при отождествлении аллофонов используются критерии простоты, симметрии и другие подобные <309, 109>.

немы, а в 1947 году — как двойные. Ч. Хоккет в 1944 году анализировал китайские придыхательные взрывные как самостоятельные согласные, а в 1947 году — как сочетание с [h]. Фонологический анализ японского языка у Б. Блока в 1946 году включал фонему [q], которая была исключена им из состава японских фонем в 1950 году (196, 258).

Итак, мы рассмотрели технические приемы, с помощью которых решается задача установления элементарных единиц языка на всех уровнях анализа. Две другие экспериментальные техники — с у б с т и т у ц и я и а н а л и з п о н е п о с р е д с т в е н н о с о с т а в л я ю щ и м, — ведущие соответственно к решению второй и третьей задач (построению классов и установлению законов связи между элементами различных классов), будут рассмотрены нами в III части книги; они вполне сохранили свое значение и в том или ином виде используются в современной структурной лингвистике.

Отметим в заключение, что, хотя дескриптивистам не удалось достичь той чистоты дистрибутивной техники, которая рисовалась им в качестве идеала, из этого отнюдь не следует, что тем самым была дискредитирована идея дистрибутивного анализа. Во всех задачах лингвистической дешифровки, которые ставятся в достаточно общем виде, изучение дистрибуции элементов, правда с учетом их числовых характеристик и прежде всего частотности, остается едва ли не единственным средством, ведущим к цели¹. Неверным было бы и заключение, что американские лингвисты не создали никакой заслуживающей внимания экспериментальной техники. Их постигла неудача в разработке чисто дистрибутивного анализа, но им удалось развить другие экспериментальные процедуры обработки «полевого» материала, применимые к языкам самого различного строя. Дескриптивисты не только более точно сформулировали некоторые уже известные принципы фонологического и морфологического анализа (129, 263), но и разработали ряд новых идей, к числу которых относятся идеи, связанные с 1) компонентным анализом фонем (295), (308); 2) суперсегментными, или просодическими элементами (интонация, ударение,

¹ Ср. лингвистические задачи А. А. Зализняка (61) и опыты практической дешифровки В. В. Шеворошкина (218), (219).

тон, пауза, стык) <301>, <308>; 3) фонологическими моделями <308>; 4) компонентным анализом морфем <291>, <295>; 5) типами морфем, среди которых выделяются впервые описанные дескриптивистами разрывные, слитные и негативные морфемы <290>, <343>; 6) типами морфологических структур и морфологических способов и процессов <343>. Они применили технику дескриптивного анализа к материалам самых различных языков (например, английского, французского, русского, японского, китайского, суахили, древнееврейского), в том числе и никогда ранее не описывавшихся, и создали надежную основу для широких типологических сопоставлений. Современная структурная лингвистика унаследовала от дескриптивистов не только эти частные идеи и материалы, но и важную в методологическом отношении мысль о том, что наиболее естественной формой лингвистической модели, которая на входе получает текст, а на выходе выдает некоторые сведения о языке, является алгоритм (см. часть II, гл. 2).

Перейдем к учению копенгагенской школы структурной лингвистики — так называемой глоссематике, которая представлена такими учеными, как Л. Ельмслев, К. Тогебю, Х. Ульдалль и немногими другими. В отличие от только что рассмотренной нами дистрибутивной лингвистики глоссематика была задумана не как экспериментальная техника обработки текстов, а как универсальная лингвистическая теория. В то время как дистрибутивистика родилась и развивалась под влиянием практических потребностей описания индейских языков Америки, решающую роль в становлении глоссематики сыграли теоретические интересы ее создателя Л. Ельмслева. Один из крупнейших лингвистов современности, феноменальный полиглот и тонкий знаток классической лингвистики, Л. Ельмслев был едва ли не первым языковедом, серьезно занимавшимся математической логикой и методологией науки. Глоссематика — плод переосмысления идейного наследия классической лингвистики, и в особенности системы Ф. де Соссюра, с точки зрения некоторых формально-логических и общеметодологических научных принципов.

Л. Ельмслев и его ближайшие единомышленники, среди которых особое место занимает Х. Ульдалль <304>,

развили взгляды Ф. де Соссюра и на язык как объект лингвистики, и на лингвистику как науку о языке.

Учение Ф. де Соссюра о языке как мысли, организованной в звучащей материи, и определение языка как формы, а не как субстанции, было развернуто Л. Ельмслевом в учение о плане выражения и плане содержания <53>, <303>. Планом выражения называется внешняя сторона языка, т. е. звуковая, графическая или иная оболочка воплощаемой в нем мысли; планом содержания называется мир мысли, находящей выражение в языке.

Однако не все в плане выражения и в плане содержания принадлежит языку. Возьмем сначала план выражения, в особенности звуки. Из большого числа физически возможных звуков каждый язык использует лишь небольшую часть, причем он использует эту часть по-своему. Субстанция, материал, из которого строится план выражения, может быть одним и тем же для всех естественных языков: это все те звуки, которые могут произноситься речевым аппаратом человека. Но форма плана выражения, т. е. способы, по которым эти звуки объединяются в систему в данном языке, неповторима. В большинстве европейских языков глухие и звонкие согласные являются разными фонемами, ср. *шесть*—*жесть*, англ. *god* — «бог» — *cod* — «треска», фр. *pas* — «шаг» — *bas* — «тихий», а в языке чиппева (Висконсин и Мичиган) они образуют варианты фонем, ср. [*gi:žik*, *gi:šik*, *ki:žik*, *ki:šik*] — «небо». С другой стороны, в европейских языках различие кратких и долгих согласных не используется обычно в качестве фонологического, а в том же языке чиппева оно играет фонологическую роль, ср. [*ki : šik*] — «небо» в противоположность [*ki : šikk*] — «кедр» <253а, 102>.

Следовательно, в плане выражения имеется субстанция выражения — фонетический, графический или иной материал, который может быть общим для ряда естественных языков. На этот материал накладывается форма выражения, или способ его использования в данном языке.

Субстанцию и форму мы находим и в плане содержания. Субстанцией содержания является все, что может быть предметом мысли, а формой содержания — способ упорядочения и комбинации идей, характерный для данного языка. Субстанция содержания одна и та же

для всех языков, а форма содержания у каждого языка своя собственная и неповторимая. Л. Ельмслев усматривает здесь полный параллелизм с планом выражения (ср. <90>).

Поясним сказанное на примере. Все люди могут мыслить понятие числа, но люди, говорящие на русском, английском, французском или немецком, оформляют свою мысль в противопоставлении единственного числа множественному, а люди, говорившие на санскрите или старославянском, оформляли свою мысль в системе с третьим членом — двойственным числом. Все люди могут мыслить понятие времени, но древние германцы оформляли свою мысль о времени только в противопоставлении прошедшего и непрошедшего, в то время как в современных германских языках прошедшее и настоящее противопоставляются не только друг другу, но и будущему времени. Это не значит, конечно, что носители русского языка не могут мыслить двойственности рядом с единичностью, а древние германцы не могли мыслить будущего. Это значит лишь, что один и тот же участок субстанции содержания по-разному систематизируется и упорядочивается в разных языках. Тот или иной способ упорядочения и создает своеобразную форму содержания из одной и той же субстанции (ср. мысли Э. Сэпира о «модели» и «покрое» языка).

Между формой выражения и формой содержания имеется определенная связь. Она определяется так называемым принципом коммутации, который можно сформулировать следующим образом: если различие в плане содержания соответствует какому-либо различию в плане выражения, то оно существенно для данного языка; с другой стороны, если различие в плане выражения соответствует какому-то различию в плане содержания, то оно также существенно для данного языка.

Вообще говоря, в плане выражения могут существовать различия, не связанные ни с какими различиями в плане содержания. Так, в русском языке существует несколько вариантов фонемы [e], ср. уже приводившийся пример *семь, сел, шесть, шест*. Они не могут рассматриваться как самостоятельные фонемы, потому что не коммутируют ни с какими различиями в плане содержания. С другой стороны, в плане содержания могут наблюдаться

различия, не связанные ни с какими различиями в плане выражения. В финском, венгерском и китайском языках одно и то же слово значит и «он» и «она» (фин. *hän*, венгер. *ő*, кит. *thā*); это различие в плане содержания никак не выражается и, следовательно, не является существенным, в отличие от таких, например, языков, как русский, английский, немецкий, французский, где оно коммутирует с различиями в плане выражения, ср. *он* и *она*, *he* и *she*, *il* и *elle*, *er* и *sie*. Аналогичным образом различие между *мыть* и *стирать*, существенное в русском языке, несущественно в английском, где одна неделимая единица — глагол *to wash* — соответствует двум разным единицам русского языка <54>.

Различия в одном плане, не коммутирующие с различиями в другом плане, образуют варианты одной и той же единицы (ср. варианты фонемы [e] в русском языке). Варианты находятся в отношении субституции друг к другу и сохраняют некоторый инвариант при всех возможных изменениях. Инварианты плана выражения Л. Ельмслев называет кенемами, а инварианты плана содержания — плеремами¹.

Язык — это система инвариантов. Следовательно, в язык в собственном смысле слова входит только форма выражения, изучаемая кенематикой, и форма содержания, изучаемая плерематикой. Субстанция выражения и субстанция содержания не входят в язык. Они являются предметом фонетики и семантики соответственно; последние не являются частями глоссематики, но выступают по отношению к ней в качестве вспомогательных дисциплин.

Все сказанное можно обобщить в следующей схеме (см. схему 1).

Задача глоссематики состоит в том, чтобы дать анализ объекта. Непосредственным объектом, к которому прилагается аппарат глоссематики, является текст; из него в результате анализа извлекается система. Предварительным этапом анализа является катализ текста, под которым понимается дополнение усеченных

¹ «Кенемы» («кенематика») соответствуют фонемам (фонологии), а «плеремы» («плерематика») — значениям (семантике). В дальнейшем везде, где это возможно без ущерба для существа дела, мы будем использовать традиционную терминологию, так как терминологические новшества Л. Ельмслева крайне непопулярны за пределами глоссематики.

план выражения		план содержания	
субстанция	форма	форма	субстанция
фонетика	кенематика	плерематика	семантика
г л о с с е м а т и к а			

или как-то поврежденных фраз до их нормальной формы. Так, русская фраза *Мы тут обменялись и не сочли* не может стать объектом анализа до тех пор, пока в ней не будут восстановлены недостающие зависимые элементы: *Мы тут обменялись мнениями и не сочли это нужным*. На первом этапе анализа текста линия выражения отделяется от линии содержания, а на последующих этапах каждая из этих линий членится на все более мелкие части. Членение производится с учетом границ между знаками (абзацами, сложными предложениями, простыми предложениями и т. д.) до тех пор, пока мы не дойдем до означающих и означаемых, соответствующих кратчайшим знакам — морфемам (точнее, морфам). Означающие и означаемые, соответствующие морфемам, членятся на элементарные единицы плана выражения и плана содержания (см. ниже), которые не являются уже означающими или означаемыми и, следовательно, не представляют собой знаков. Переход от знаков к незнакам знаменует переход от текста к системе.

Каждый раз, когда в ходе анализа проводится то или иное деление, устанавливается характер зависимости или функции между выделяемыми частями, независимо от того, являются ли они единицами текста или элементами системы.

Рассматриваются три общих типа глоссематических функций, под которые подводятся отношения между любыми единицами плана выражения и плана содержания и в тексте, и в системе:

1) Двусторонняя зависимость, или интердепенденция, имеющая место между двумя элементами, не существующими один без другого.

Таково отношение между гласными и согласными, существительными и глаголами (в системе); между подлежащим и сказуемым, морфемой падежа и морфемой числа в латинских или русских существительных (в тексте).

2) **Односторонняя зависимость**, или **детерминация**, имеющая место между двумя элементами, один из которых предполагает другой. Таково отношение между любым косвенным падежом, например дательным и именительным: если в системе есть дательный, в ней обязательно должен быть и именительный; обратное неверно. Примером детерминации в тексте являются некоторые случаи управления (так, предлоги *для*, *возле*, *у* предполагают родительный падеж, но обратное неверно); отношение между суффиксом и основой в производном слове (суффикс предполагает основу, но не наоборот); отношение между согласным и гласным в слоге (согласный предполагает гласный, но не наоборот).

3) **Свободная зависимость**, или **констелляция**, имеющая место между двумя элементами, каждый из которых может существовать без другого. Таковы отношения между категорией лица и рода русского глагола в системе (в настоящем времени выражается лицо и не выражается род, а в прошедшем времени выражается род и не выражается лицо). Примером констелляции в тексте являются отношения между латинским предлогом *ab* и аблативом или русским предлогом *в* и винительным падежом: предлоги *ab* и *в* управляют и другими падежами, а аблатив и винительный падеж управляются и другими предлогами.

Положения, касающиеся плана выражения и плана содержания, включая принцип коммутационной проверки и понятие вариантов и инвариантов, выводимы непосредственно из учения Ф. де Соссюра о языке и речи, языковом знаке и значимости. В отличие от этого, учение о трех типах глоссематических функций является гораздо более оригинальным, и хотя в наше время мало кто пользуется именно этими функциями, Л. Ельмслеву все же принадлежит та несомненная идейная заслуга, что он практически первым из лингвистов сделал сознательную попытку найти понятия, достаточно общие для того, чтобы с их помощью можно было описывать отношения между любыми языковыми единицами любых уровней в тексте (фонемами в составе слога, морфемами в составе

слова, словами в составе предложения) и в системе (элементами в составе класса, классом и элементом, классами в составе объемлющего множества).

Л. Ельмслев и его последователи были пионерами в еще одной важной области — разработке учения о знаках и фигурах. Извлекая из текста систему, мы сводим бесконечные инвентари единиц, встречающихся в тексте, к конечному инвентарю единиц, составляющих язык. Средством, с помощью которого бесконечные инвентари сводятся к конечным и даже ограниченным инвентарям, является разложение знаков на фигуры. Фигура, в отличие от знака, является односторонней единицей. У фигур плана выражения (фонем) нет означаемых, а у фигур плана содержания (смысловых категорий) нет означающих. Число знаков бесконечно или очень велико (ср. абзацы, предложения, слова, морфемы), а число фигур весьма ограничено. Так, в любом языке число фонем не превосходит 70—80, а число фигур содержания, к которым Л. Ельмслев относит и все грамматические категории (число, падеж, лицо, время и т. д.), не намного больше. «Таким образом, язык организован так, что с помощью горстки фигур и благодаря их все новым и новым расположениям (лучше сказать, «сочетаниям» или «комбинациям». — Ю. А.) может быть построен легион знаков. Если бы язык не был таковым, он был бы орудием, неподходящим для своей задачи» (53, 305) ¹.

На учении о фигурах содержания стоит задержаться несколько подробнее, так как оно, вместе с некоторыми идеями Э. Сэпира (56), Р. О. Якобсона (317) и некоторых других исследователей (291), явилось прообразом современных представлений о структурной семантике (см. работы о «семантических множителях» в части IV). По словам Л. Ельмслева, в традиционной грамматике «не было сделано даже попытки предпринять такое разложение знакового содержания, хотя соответствующее разложение знакового выражения на фигуры выражения так же старо, как и изобретение фонетического письма... Данное несоответствие имело чрезвычайно катастрофические последствия: встречаясь с неограниченным числом

¹ В указанном свойстве Л. Ельмслев видит «наиболее существенную черту в структуре любого языка» (53, 305). В связи с этим он фактически порывает с соссюрровским представлением о языке как чисто знаковой системе.

знаков, люди считали анализ содержания неразрешимой проблемой...» (53, 325). Между тем с точки зрения глоссематики процедура здесь точно такая же, как и в случае с фигурами плана выражения. Фигуры содержания выделяются из минимальных семантических противопоставлений; так, в паре *девочка* — *мальчик* выделяется фигура (1) «мужской — женский» (пол), в паре *девочка* — *женщина* выделяется фигура (2) «юный — зрелый» (возраст), а в паре *девочка* — *тёлка* выделяется фигура (3) «человеческий — животный» (вид). Фигура (1) составляет часть содержания слов *женщина, девушка, мать, корова, кобыла, овца, курица* и слов *мужчина, юноша, отец, бык, жеребец, баран, петух*; фигура (2) составляет часть содержания слов *мальчик, девочка, ребенок, цыпленок, ягненок, жеребенок* и слов *мужчина, женщина, взрослый, петух, лошадь, жеребец* и т. д. Позднее для записи значений Х. Ульдалль (304) применил несложную математическую символику: фигуры, входящие в содержание знака, он объединил знаком математической конъюнкции, т. е. знаком связи, которой соответствует союз *и*. Значение слова *девочка* он бы записал так: «женский · юный · человеческий». На этой основе он впервые стал составлять семантические равенства типа *мальчик* = «мужской · юный · человеческий», откуда можно извлечь $\frac{\text{мальчик}}{\text{мужской}} = \text{«юный · человеческий»} = \text{«ребенок»}$. Этот новый гибкий подход к значениям открыл перед лингвистикой возможности, вполне оцененные только много лет спустя.

Итак, мы познакомились со взглядами глоссематиков на язык как объект лингвистики и на задачи лингвистического анализа. Теперь нам остается рассмотреть их концепцию лингвистической теории.

Теория должна дать нам понимание нашего объекта, т. е. текста, причем не только того, который послужил непосредственным поводом для создания теории, но и всех других фактически встретившихся или принципиально мыслимых, т. е. «приемлемых для носителя данного языка» (53, 33), текстов, включая и те тексты, которые не только реально не существуют, но, вероятно, никогда не будут существовать. Более того, «на основе информации, которую она (теория) дает о языке вообще, она должна быть полезна для описания и предсказания любого возможного текста на любом языке» (53, 277). Изучив

ограниченную выборку текстов на ограниченном числе языков, «лингвист-теоретик... должен предвидеть все мыслимые возможности — представить эти возможности, которые он сам не испытал и не видел, реализованными... Только таким образом можно создать лингвистическую теорию, которую с уверенностью можно применять» <53, 277>.

Из сказанного следует, что лингвистическая теория не может строиться индуктивно. Пока мы пытаемся только обобщить непосредственно наблюдаемые факты, у нас нет никакой гарантии, что наши обобщения не будут разрушены при некотором увеличении материала. Но, поскольку весь материал, ввиду его бесконечности, охватить невозможно, индуктивно построенная эмпирическая система никогда не сможет претендовать на общеприменимость своих положений: она всегда будет «обобщением» для данного материала, имеющим только частный интерес. Выход из этих затруднений только один — строить лингвистическую теорию не индуктивно, а дедуктивно. Лингвистическая теория должна иметь вид математического исчисления «всех мыслимых возможностей в определенных рамках» <53, 278>.

Когда теория, в которой исчислены все логические возможности, заложенные в данной системе понятий, построена, описание конкретного материала, т. е. грамматики конкретного языка, получается «наложением» теории на язык и указанием того, какие из предусмотренных в ней возможностей реализуются в данном случае. Теория обладает свойством универсальности: ее понятия настолько общи, что они могут быть приложены к материалу любого языка. В этом отношении интересна работа Л. Ельмслева о падежах <302>, в которой исчислены все возможные падежные противопоставления. Значение любого падежа любого языка может быть представлено как комбинация значений трех абстрактных признаков: 1) направления, 2) контакта, 3) субъективности. Каждый признак может принимать 6 различных значений (ср. значения приближения, удаления, покоя и т. п. для признака «направление»). Признаки логически независимы друг от друга, и поэтому всего возможно $6 \cdot 6 \cdot 6 = 216$ различных комбинаций значений признаков. Следовательно, максимальное число падежей в системе равно 216, а минимальное — 2. Падежные системы всех реально суще-

ствующих языков располагаются между этими двумя полюсами. Так, применение исчисления к материалу английского языка дает четыре падежа: субъектный, родительный, дательный, транслятивный, ср. *The boy* (субъектный) *sent his mother* (дательный) *a letter* (транслятивный) — «Мальчик послал своей матери письмо». Теория, таким образом, может служить языком-посредником для типологического сопоставления языков: сделанное на ее основе описание конкретного языка сравнимо с описанием любого другого языка, построенным на той же базе.

Поскольку глоссематика не учитывает материала, из которого построены ее объекты, рассматривая каждый объект как точку пересечения некоторых функций, ее аппарат, по мнению Л. Ельмслева, применим к любой знаковой системе, форма которой аналогична форме естественного языка. Глоссематика предстает, таким образом, как общая теория любых знаковых систем, «хотя по замыслу ее главная цель состояла в установлении основы для описания лингвистического и иного гуманитарного материала» <304, 86>.

Нам остается рассмотреть в заключение взгляды глоссематиков на связь их теории и реальности. С точки зрения Л. Ельмслева, теория сама по себе, будучи чисто дедуктивной системой, «ничего не говорит ни о возможности ее применения, ни об отношении к опытным данным» <53, 274>. В этом смысле теория является произвольной, и единственное требование, которому она должна удовлетворять, исчерпывается сформулированным Л. Ельмсловом принципом эмпиризма (внутреннего совершенства): «Описание должно быть свободным от противоречий, исчерпывающим и предельно простым» <53, 272>. Однако это свойство теории представляет лишь одну сторону вопроса. «С другой стороны, теория включает ряд предпосылок, о которых из предшествующего опыта известно, что они удовлетворяют условиям применения к некоторым опытным данным» <53, 275>. Хотя в самую теорию никак не входит утверждение о том, что ее объект существует, она во всех своих применениях «должна обнаружить результаты, согласующиеся с так называемыми (действительными или предполагаемыми) экспериментальными данными» <53, 271>. Это — критерий внешней оправданности, пригодности или применимости

теории, и если она ему не удовлетворяет, она должна быть отброшена, несмотря на свое формальное совершенство.

К сожалению, фактически теория Л. Ельмслева почти не применялась для практического описания естественных языков или других семиотических систем, если не считать нескольких работ Л. Ельмслева и одной, по общему признанию неудачной, книги К. Тогебю <367>, и вызвала этим критическую реакцию лингвистов самых различных направлений. А. Мартине пошел настолько далеко, что назвал ее «башней из слоновой кости» <105, 51>.

Оппоненты Л. Ельмслева правы лишь отчасти. Нельзя не видеть, что глоссематика во многих отношениях опередила современную ей науку и оставила заметный след в истории лингвистической мысли. Л. Ельмслев яснее, чем кто бы то ни было до него, понял принципиальное значение дедуктивных методов для будущего развития лингвистики. Правда, созданное Л. Ельмсловом и Х. Ульдаллем исчисление, представляющее собой систему определений, которые управляются лишь простейшими принципами комбинаторики, не стало основой такого развития, но оно было первым шагом на верно выбранном пути. В этом отношении особый интерес представляют следующие идеи Л. Ельмслева: 1) определения лингвистических понятий должны строиться таким образом, чтобы последующие определения выводились из предыдущих и вся система сводилась к небольшому числу исходных (неопределяемых) понятий самого общего характера, не нуждающихся в обосновании¹ (106 определенных Л. Ельмсловом глоссематических терминов сводимы к 4—5 исходным понятиям типа «функция», «тождество» и т. п.); 2) наиболее естественной формой для дедуктивной лингвистической модели является исчисление; 3) лингвистическая теория должна включаться в общую теорию семиотических систем, по отношению к которой конкретные теории тех или иных предметных областей (например, грамматики различных языков) выступают в качестве объектов. Современная лингвистика унаследовала и некоторые другие введенные Л. Ельмсловом принципы и понятия, даже если они

¹ Ср. понятие «дефиниционной» лингвистической модели у П. Гарвина <276>, <278>.

не сохранили своих глоссематических наименований. Таковы понятия коммутации, катализа и фигуры содержания.

На этом мы закончим характеристику глоссематики и перейдем к пражской школе структурализма — функциональной лингвистике¹.

Если американская дескриптивная лингвистика представляет собой «предписание об описании», а копенгагенская глоссематика — общую семиотическую теорию, то пражская функциональная лингвистика является (в идеале) теорией реальных явлений и процессов естественного языка. В этом она сходна с традиционной описательной грамматикой и поэтому производит, по сравнению с другими школами структурной лингвистики, несколько обманчивое впечатление традиционности.

Расцвет деятельности Пражского лингвистического кружка приходится на 30-е годы; в 40-е годы его творческая активность резко упала, а к 1953 году он распался и организационно <28>². Тем не менее почти все собственно лингвистические идеи пражского структурализма разрабатываются и сейчас, причем это делается на более широкой, чем прежде, теоретической и технической основе. В той или иной мере влияние идей пражцев сказывается на работе в области современной лингвистики, проводимой в Советском Союзе <40>, <136>, ГДР <80>, <315>, Чехословакии <371>, Румынии <333>, США <235>, <236> и других странах.

Школу функциональной лингвистики представляют чешские языковеды В. Матезиус, Б. Трнка, Б. Гавранек, И. Вахек, В. Скаличка и русские языковеды Н. С. Трубецкой, Р. О. Jakobson и С. Карцевский, долгое время работавшие в Праге. Хотя никто из них не создал³ обоб-

¹ Превосходный очерк функциональной лингвистики, написанный Т. В. Булыгиной, читатель найдет в уже упоминавшейся монографии <28>.

² Впрочем, в последнее время появились признаки того, что кружок будет возрожден, хотя, может быть, и в другой форме. Об этом свидетельствует, в частности, выход в свет в 1964 году первого тома «Пражских лингвистических трудов» <371>.

³ Сейчас это уже неверно; см.: Josef Vachek, *The linguistic school of Prague. An introduction to its theory and practice*. Bloomington and London, 1966.

щающего труда, посвященного основам функциональной лингвистики, их учение представляет собой довольно стройную систему взглядов. Оно выросло из сформулированного еще в 1929 году в знаменитых «Тезисах Пражского лингвистического кружка» положения «о языке как функциональной системе», т. е. «системе средств выражения, служащей какой-то определенной цели» <176, 69>. Свою задачу пражские языковеды видят в том, чтобы обнаружить эту систему во всех аспектах языка — фонологическом, морфологическом, синтаксическом и даже лексическом ¹.

Наиболее полно пражцы разработали фонологическое учение, систематически изложенное в «Основах фонологии» Н. С. Трубецкого ². Этот энциклопедический труд, содержащий описание около 200 фонологических систем различных языков, подводит итоги более чем десятилетней теоретической деятельности Н. С. Трубецкого и в значительной мере всего пражского направления. Поэтому он будет взят нами за основу при изложении фонологических воззрений пражцев.

В полном соответствии с обычным структурным представлением о языке Н. С. Трубецкой разграничивает фонетику и фонологию: «Фонология так относится к фонетике, как политическая экономия к товароведению или наука о финансах к нумизматике» <179, 18>. Причисляя только фонологию к собственно лингвистическим дисциплинам, Н. С. Трубецкой тем не менее считает возможным использовать при анализе функций фонологических единиц фонетические данные. Он рассматривает три основные функции фонологических элементов: 1) вершинную, или кульминативную, состоящую в указании того, «какое количество «единиц» (=слов, словосочетаний) содержится в данном предложении» <179, 36>; 2) разграничительную, или делимитативную, состоящую в указании границ между единицами; 3) смысловозначительную, или дистинктивную, состоящую в различении значащих единиц. Учение о смысловозначении является ядром фонологической концепции Н. С. Тру-

¹ Впрочем, программа лексико-семантических исследований не была ими реализована.

² См. об этой работе <157>.

бецкого и других пражских структуралистов и поэтому заслуживает более внимательного анализа.

Центральным понятием этого учения является понятие фонологической, или смысловозначительной, оппозиции, под которой понимается «звуковое противоположение», способное «дифференцировать значение двух слов данного языка» <179, 38>, ср. русское *роль* — *моль*, фр. *sein* («грудь») — *son* («звук»), нем. *stillen* («успокаивать») — *Stollen* («штольная»), англ. *line* («линия») — *lane* («дорожка»). Понятие фонемы является производным от понятия фонологической оппозиции: фонема определяется как совокупность фонологически существенных признаков. Этим фонологическая концепция пражских структуралистов отличается от соответствующих концепций дескриптивистов и глоссематиков: дескриптивисты, за редкими исключениями <308>, рассматривали «компонентный анализ» фонем как нечто второстепенное, а Л. Ельмслев считал фонему элементарной фигурой плана выражения и вообще отрицал целесообразность ее разложения на компоненты. Можно лишь сожалеть о том, что ни дескриптивисты, ни глоссематики не оценили по достоинству этого восходящего к И. А. Бодуэну де Куртенэ (см. стр. 25) представления о фонеме как «пучке различительных признаков»: оно революционизировало фонологию и имело далеко идущие последствия (см. ниже).

Изложенное выше определение фонемы не является конструктивным; оно не содержит никаких указаний о том, как практически выделить фонемы некоторого языка. В связи с этим Н. С. Трубецкой формулирует технические правила идентификации фонем, по надежности и богатству указаний превосходящие все, что сделано в этой области дескриптивистами и глоссематиками. Для выделения фонем языка необходимо уметь отличать 1) фонему от ее вариантов (парадигматическая идентификация фонем) и 2) фонему от многофонемного сочетания (синтагматическая идентификация фонем). Правила парадигматической идентификации фонем, сформулированные значительно раньше соответствующих предписаний дистрибутивистов, во многом аналогичны им; в частности, факультативные и комбинаторные варианты фонем выделяются Н. С. Трубецким в условиях, соответствующих свободному чере-

дованию и дополнительной дистрибуции, а разные фонемы — в условиях, соответствующих контрастной дистрибуции (для уяснения этих правил читатель может обратиться к примерам, приводимым нами на стр. 49—50). Что касается правил синтагматической идентификации фонем, то они не имеют каких-либо аналогов в дескриптивистике или глоссематике. Реализацией одной фонемы считается: 1) сочетание звуков, составные части которого в данном языке не распределяются по двум слогам; по этому правилу группа звуков [ts] представляет собой одну фонему в русском (ср. *лицо*, *целый*), чешском и польском (ср. *co* — «что»), но две фонемы в финском, где эта группа возможна только на стыке двух слогов (ср. *it-se* — «сам»); 2) сочетание звуков, образуемое с помощью единой артикуляции, ср. английские дифтонги [ai], [ei], [oi]; 3) сочетание звуков, длительность которого не превышает длительности других фонем данного языка, и т. д. Всего Н. С. Трубецкой предложил семь правил синтагматической идентификации фонем, причем в большинстве из них решающую роль играют фонетические критерии. У некоторых исследователей эта «нефонологичность» Н. С. Трубецкого вызвала возражения, по всей видимости недостаточно обоснованные, так как Н. С. Трубецкой смотрел на свои правила как на чисто практические процедуры и не включал их в теорию¹.

Поскольку фонемы определяются через фонологические признаки, последние должны считаться элементарными и основными единицами фонологической системы языка. «В фонологии основная роль принадлежит не фонемам, а смысловозначительным оппозициям. Любая фонема обладает определенным фонологическим содержанием лишь постольку, поскольку система фонологических оппозиций обнаруживает определенный порядок или структуру. Чтобы понять эту структуру, необходимо исследовать различные типы фонологических оппозиций» (179, 75).

Подробное изложение единственной в своем роде классификации фонологических оппозиций, предложенной

¹ Ср. постановку вопроса о парадигматической и синтагматической идентификации («делимитации») фонем у К. Л. Эбелинга <271> и С. К. Шаумяна <205>.

Н. С. Трубецким, не входит в наши задачи. Мы остановимся ниже лишь на некоторых наиболее интересных ее деталях.

Оппозиции классифицируются по трем различным признакам: 1) по отношению данной оппозиции ко всей системе оппозиций; 2) по отношению между членами оппозиции; 3) по объему смысловозначительной силы оппозиции.

Наиболее важными для наших целей являются две последние классификации.

С точки зрения отношений между членами оппозиции выделяются а) п р и в а т и в н ы е оппозиции, образующиеся на основе бинарных признаков типа «звонкость» — «незвонкость», «назализованность» — «неназализованность», или, в общем виде, A — не A ; ср. $b—p$, $\bar{a}—a$ и т. п.; б) г р а д у а л ь н ы е, или с т у п е н ч а т ы е, оппозиции, члены которых различаются не наличием — отсутствием признака, а его степенью, ср. $i—e—e$; в) э к в и п о л е н т н ы е, или р а в н о з н а ч н ы е, оппозиции — все остальные; ср. $p—t$, $f—k$ и т. п.

По объему различительной силы оппозиции делятся на п о с т о я н н ы е и н е й т р а л и з у е м ы е. Две фонемы образуют постоянную оппозицию, если они различаются во всех возможных положениях, ср. фр. [a] — [o]; две фонемы образуют нейтрализуемую оппозицию, если они различаются в некоторых положениях и не различаются в других, ср. рус. [t] и [д] в начале и конце слова: *том* — *дом* в противоположность *рот* — *род*. Позиция, в которой происходит нейтрализация противопоставлений, называется п о з и ц и е й н е й т р а л и з а ц и и; в этой позиции реализуется а р х и ф о н е м а, под которой понимается «совокупность различительных признаков, общих для двух фонем» (179, 87). Архифонема может совпадать с одним из членов оппозиции или быть промежуточной между ними.

С понятием привативной нейтрализуемой оппозиции у Н. С. Трубецкого связывается понятие м а р к и р о в а н н о г о и н е м а р к и р о в а н н о г о ч л е н а оппозиции, перенесенное позднее в морфологию и распространенное далеко за пределами функциональной лингвистики. Под маркированным членом оппозиции понимается тот ее член, который характеризуется наличием некоторого признака, а под немаркированным — тот член,

который характеризуется его отсутствием. Считается, что «беспризнаковым» является тот член оппозиции, который совпадает с архифонемой в позиции нейтрализации. Таковы, например, глухие согласные по отношению к звонким в русском языке: противопоставление глухих звонким нейтрализуется в конце слова, причем архифонема совпадает с глухим согласным; следовательно, именно он является беспризнаковым, немаркированным членом оппозиции. В связи с этим он определяется как «архифонема + 0» (нуль), а его маркированный коррелят как «архифонема + некоторый признак»¹.

Фонологическое учение пражцев последовательно развивается в современной структурной лингвистике Р. О. Якобсоном, который в свое время был одним из вождей пражского структурализма. Ему и его коллегам принадлежит серия экспериментальных и теоретических работ по дихотомической фонологии <235>, <236>, в которых была сделана успешная попытка трактовать все типы оппозиций как бинарные привативные оппозиции². В отличие от Н. С. Трубецкого, он рассматривает не артикуляторные, а акустические признаки фонем, причем каждый акустический признак фиксируется экспериментально (и визуально) с помощью специальной спектрографической аппаратуры.

Возвращение к акустическим признакам и использование системы бинарных противопоставлений привели к двум важным результатам:

1) В то время как для описания гласных и согласных в терминах их артикуляций нужны в принципе две разные системы артикуляторных признаков, для описания гласных и согласных на акустической основе достаточно одной системы акустических признаков. Она оказывается общей не только для гласных и согласных, но и для фонологических систем разных языков. Это позволило Р. О. Якобсону и его коллегам построить универсальную

¹ Корректнее было бы говорить не о наличии и отсутствии признака, а о ненулевом и нулевом значении признака. См. также критику фонологического понятия маркированности — немаркированности у А. А. Реформатского <158, 112—115> и С. К. Шаумяна <205, 138—142>.

² Интерпретация всех фонологических оппозиций как бинарных была сочувственно встречена некоторыми учеными <205>, но вызвала возражения у других <158>, <104>.

систему из 12 бинарных дифференциальных признаков, годную для описания фонологической системы практически любого языка. 2) На акустической основе удалось установить корреляции пар фонем, до того не замеченные и интуитивно не очевидные, ср. $k : p = a : \text{э}$ (противопоставление компактного и диффузного в обеих парах) или $p : t = u : i$ (противопоставление низкого и высокого в обеих парах). Этот факт с методологической точки зрения имеет неопределимое значение. До сих пор в лингвистике стремились к тому, чтобы теоретические выводы подтверждались интуитивными представлениями среднего носителя языка о том или ином явлении. Это требование соответствует, в общем, начальному этапу развития науки, когда интуиция исследователя не отличается сколько-нибудь серьезно от интуиции «человека с улицы». В экспериментах Р. О. Jakobsona и его коллег лингвистика впервые заглянула в такие глубины объекта, о которых средний носитель языка не подозревает, и добыла экспериментальные данные, опровергающие его неразвитую интуицию, хотя, может быть, и угаданные более тонкой интуицией исследователя, изолированной в экспериментальной работе.

Фонологическое учение пражцев развивалось и в других направлениях. Еще в 30-е годы оно было с успехом применено к решению некоторых проблем диахронической фонологии <28, 84—87>, <104>, которая фактически была создана усилиями языковедов, входивших в Пражский лингвистический кружок или близких к нему. Заслуживает упоминания предпринятая в недавнее время формализация учения Н. С. Трубецкого на основе математической теории множеств <333>. Однако для нас наибольший интерес представляет попытка В. Скалички, Р. О. Jakobsona, Б. Трнки и др. <28>, <316>, <317>, <366>, <371> перенести систему понятий, разработанных первоначально на фонологическом материале, в область морфологии¹.

¹ С этой точки зрения важны работы А. де Гроота <282>, Е. Куриловича <92>, И. И. Ревзина <150>, Е. В. Падучевой <131>, А. В. Исаченко <79>, <80>, М. В. Панова <136>, <138> и некоторых других. В них делается попытка интерпретировать на морфологическом и синтаксическом материале понятия позиции, оппозиции и нейтрализации оппозиций.

Взгляды пражских структуралистов по основным вопросам морфологии можно суммировать следующим образом:

1) Основной единицей морфологической системы языка является м о р ф е м а; всякую морфему можно представить как пучок элементарных м о р ф о л о г и ч е с к и х о п п о з и ц и й, например оппозиции падежей, оппозиции глагольных форм времени и т. п. «Репертуар» семантических признаков, служащих основой для этих оппозиций, выделяется способом, вполне аналогичным тому, с помощью которого выделяются фонологические признаки в фонологии. Так, русское глагольное окончание *-ем* разлагается на элементарные семантические значения (сема) числа (в противопоставлении *у-(ю)*), лица (в противопоставлении *-ете*), времени (в противопоставлении *-ли*) и т. д. Аналогичные взгляды на морфему как совокупность «компонентов» или «фигур содержания» развивали З. Харрис <295> и Л. Ельмслев <53>.

2) В ряде позиций морфологические противопоставления могут нейтрализоваться; так, в русском языке в неодушевленных именах существительных мужского рода нейтрализуется противопоставление именительного и винительного падежей; в английском языке в прошедшем времени глаголов нейтрализуются противопоставления по лицам и числам; в немецком языке во множественном числе нейтрализуется противопоставление по роду и т. д. Позднее в американской дескриптивной лингвистике <352> было предложено дополнить понятие нейтрализации морфологических противопоставлений понятием архиморфемы (по аналогии с архифонемой).

3) Морфологические оппозиции бинарны. С этой точки зрения русская падежная система описывается не на основе одного шестизначного (или восьмизначного) признака, как это делается в традиционной грамматике, а на основе трех двузначных: (1) периферийность (дательный, творительный, предложный) — непериферийность (именительный, родительный, винительный); (2) направленность (дательный, винительный) — ненаправленность (именительный, творительный); (3) объемность (родительный, предложный) — необъемность (именительный, дательный, винительный, творительный). Признак направленности-ненаправленности к родительному и предложному падежам неприменим. Эта разработанная Р. О. Якоб-

соном система позволяет описать не только шестипадежный, но и восьмипадежный вариант русской системы падежей <317>.

4) Морфологические оппозиции асимметричны <316>: один из членов морфологической оппозиции является маркированным (сильным) членом, а другой — немаркированным (слабым). Понятие маркированности было, как нам кажется, существенно уточнено Р. О. Якобсоном: он предложил считать маркированным тот член морфологической оппозиции, в котором выражено некоторое значение *А*, а немаркированным — тот член оппозиции, который оставляет это значение невыраженным. Таким образом, он заменил противопоставление «*А — не А*» или «*выражено А — выражено не А*» противопоставлением «*выражено А — не выражено А*». Так, в словах *учительница, ткачиха, поэтесса, секретарша* (примеры взяты из <136, 24>) выражено значение «лицо женского пола, охарактеризованное по профессии», а в словах *учитель, ткач, поэт, секретарь* это значение остается невыраженным. Аналогичным образом в видовых парах глагол совершенного вида указывает, по характеристике С. Карцевского <324>, на доведение действия до качественного предела и является, таким образом, маркированным членом противопоставления; глагол несовершенного вида не выражает этого значения и определяется, следовательно, как немаркированный член противопоставления. Маркированный член противопоставления имеет более узкую по сравнению с немаркированным членом сферу употребления; так, во всех случаях, когда употребляются слова *учительница, ткачиха, поэтесса*, можно употребить и слова *учитель, ткач, поэт*, но обратное неверно¹.

По сравнению с фонологией и морфологией синтаксическое учение пражцев, разработанное прежде всего В. Матезиусом <337>, носит менее законченный харак-

¹ Кажется естественным связать термин «сфера употребления» со статистическим понятием частотности: широта сферы употребления некоторой единицы находится в прямой зависимости от ее частотности. Такое предположение подтверждается как будто и тем фактом, что часто употребляемые единицы беднее семантическими признаками, чем редко употребляемые. Оказывается, однако, что в общем случае статистическая интерпретация понятия «сферы употребления» неверна. Так, маркированными членами видовых пар считаются обычно глаголы совершенного вида, хотя в текстах они встречаются, по данным частотного словаря Х. Йоссельсона <322>, значительно чаще, чем глаголы несовершенного вида.

тер. Однако даже в этой области, объективно менее подготовленной для широких обобщений, чем фонология и морфология, пражцам удалось сформулировать ряд плодотворных идей, воспринятых и развиваемых современной лингвистикой. Это относится в первую очередь к ключевому положению синтаксического учения пражцев — противопоставлению «формального членения» предложения на грамматическое подлежащее и сказуемое «актуальному членению»¹ предложения на «новое» и «данное», «рему» и «тему». «Новое» (в обычной терминологии — психологический, или логический, предикат) — это то, ради чего делается сообщение, а «данное» (в обычной терминологии — психологический, или логический, субъект) — это то, о чем делается сообщение. При этом существенно то, что грамматическое подлежащее не всегда совпадает с «данным», а грамматическое сказуемое — с «новым». В предложении *Птица летит* такое совпадение имеет место, а в предложении *Летит птица* «данным» является грамматическое сказуемое, а «новым» — грамматическое подлежащее. Простота этого примера не должна быть истолкована превратно. Глубокий смысл учения об «актуальном членении» состоит в том, что оно привлекает внимание к чрезвычайно важным в структуре естественных языков механизмам передачи логического акцента (логического ударения, подчеркивания), роль которого была вполне оценена в результате развития работ по семантическому машинному переводу <110а>. Хотя, как справедливо замечает В. В. Виноградов, некоторые положения учения об «актуальном членении» не выходят за пределы традиционных представлений <35, 24—25>, пражцам принадлежит та бесспорная заслуга, что в традиционной синтаксической проблематике они сумели увидеть и выделить в качестве центрального комплекс вопросов, который действительно имеет первостепенный интерес для лингвистической теории.

В заключение упомянем о типологических исследованиях пражцев. Серьезная типология может строиться либо на основе общей дедуктивной теории языка, либо на

¹ Лучше было бы говорить не об «актуальном членении» предложения, а о «смысловом», или «логическом». Термин «актуальное членение» не передает фактического содержания понятия и не воспринимается как противопоставленный термину «формальное («грамматическое») членение».

основе языковых универсалий. Если Л. Ёльмслев с его идеей лингвистики как общей теории знаковых систем пытался реализовать первую возможность, то пражцы с их интересом к грамматическим явлениям конкретных языков естественно выбрали второй путь — путь поисков языковых универсалий. Наряду с Э. Сэпиром и его последователями многие из них были пионерами в этой новой и интересной области, а Р. О. Якобсону мы обязаны самым солидным в истории лингвистики трудом по языковым универсалиям, написанным по его инициативе и при его живейшем участии <185>, <373>.

Таким образом, деятельность трех описанных нами школ структурной лингвистики не прошла бесследно. Они создали фонологию, написали несколько замечательных глав морфологии, реформировали диахроническую лингвистику, начали серьезную разработку типологических проблем. В этих школах были высказаны важные мысли о том, как должна строиться лингвистическая теория. Даже в тех случаях, когда их искания «не имели бесспорно положительных результатов, значение их для развития языкознания заключалось в том, что они заставляли критически пересматривать старые догмы и способствовали осознанию несовершенств старых методов» <172, 40>. По всей видимости, они сделали бы еще больше, если бы им удалось наладить сотрудничество.

К сожалению, взаимопонимание между школами было серьезно затруднено из-за того, что не было предпринято никакой попытки совместно разработать единый язык лингвистики. Языки дескриптивистики, глоссематики и функциональной лингвистики развивались совершенно независимо друг от друга. Это привело к тому, что одним и тем же объектам или принципам присваивались различные наименования, а совершенно различные вещи назывались одинаково. Так, принцип коммутиационной проверки, используемый глоссематиками для отождествления и разграничения языковых единиц, вполне сравним с дистрибутивным правилом установления тождеств и различий в условиях контрастной дистрибуции, разработанным дескриптивистами, и с критерием смысла различительных противопоставлений, который был предложен пражскими функционалистами. Дескриптивное понятие аллофонов, находящихся в дополнительной

дистрибуции, принципиально не отличается от понятия комбинаторных вариантов фонемы у Н. С. Трубецкого и понятия обусловленных вариантов фонемы, или вариаций, которым пользовался Л. Ельмслев. Deskриптивное понятие аллофонов, находящихся в свободном чередовании, принципиально не отличается от понятия факультативного варианта фонемы у Н. С. Трубецкого и понятия свободного варианта или варианта у Л. Ельмслева. Идея нейтрализации (морфологических) противопоставлений, принадлежащая пражцам, близка к идее синкретизма, развитой Л. Ельмслевом. Отношение детерминации, рассматриваемое Л. Ельмслевом, в ряде случаев сводимо к отношению маргинального элемента к ядерному, являющемуся основой дистрибутивного анализа по непосредственно составляющим.

С другой стороны, совершенно по-разному используются в трех школах термины *субституция*, *инвариант*, *функция* и некоторые другие. У американских дескриптивистов возможность *субституции* является в большинстве случаев признаком различия языковых единиц, а у Л. Ельмслева *субституция* (в отличие от коммутации) всегда является признаком их тождества. *Инвариант* в смысле Л. Блумфильда и ряда других дескриптивистов — это некоторая постоянная часть, или неизменное свойство, множества *реальных единиц*, а у Л. Ельмслева — это абстрактная, *идеальная единица*. *Функция* для пражцев — это роль, которую выполняет та или иная языковая единица, в то время как для глоссематиков — это зависимость одной единицы от другой. Положение осложняется тем, что многие из этих терминов (*функция*, *коммутация*, *дистрибуция*, *отношение*, *инвариант*) заимствованы из математики, но значение, которое в них вкладывается лингвистами, не вполне совпадает с тем, которое они имеют в соответствующих математических дисциплинах.

Неработанность языка лингвистики явилась одной из основных причин изоляционизма школ, который привел к разобщению их усилий и созданию самостоятельных лингвистических учений, внешне как будто мало

связанных друг с другом. Это дало некоторым лингвистам повод заговорить о расколе лингвистики.

С нашей точки зрения для такого вывода нет достаточных оснований. Выше (на стр. 36) мы уже отмечали, что разные школы объединяет прежде всего соссюроевское понимание языка как объекта лингвистики. Но учения различных школ не только связаны друг с другом. В некотором смысле они дополняют друг друга, потому что, помимо общего объекта — языка в соссюроевском смысле этого термина, — у них есть свои специфические объекты, изучение которых в равной мере необходимо для создания единой и цельной лингвистической теории. Мы уже говорили о том, что пражцев занимал прежде всего непосредственный объект лингвистики, т. е. тот или иной естественный язык. Несколько модернизируя их представления о задачах науки, мы могли бы сказать, что их интересовали лингвистические модели реального объекта, лингвистические модели конкретных языковых процессов. Обычной моделью такого рода является грамматика того или иного языка. Американских дескриптивистов всегда интересовала не только и не столько грамматика конкретного языка, сколько процедуры, ведущие к ее открытию. Используя современные понятия, мы могли бы сказать, что они разрабатывали универсальный алгоритм построения конкретных грамматик тех или иных языков по их текстовым данным. Наконец, Л. Ельмслев сделал попытку разработать общую теорию конкретных грамматик. Глоссематика — это, в идеале, формальная теория всех семиотических систем, объектами которой являются не реальные языки, а уже построенные грамматики (описания) реальных языков. В сущности, «эмпирический принцип» Л. Ельмслева можно считать тем аппаратом, с помощью которого оцениваются различные модели одного и того же объекта, например грамматики одного и того же языка.

Мы видим, таким образом, что в идеале, если бы дескриптивисты, глоссематики и функционалисты занимались разработкой своих учений с учетом того, что делают их коллеги, результаты их деятельности дополняли бы друг друга, образуя в совокупности все здание лингвистической теории. Отметим, что три названных выше типа моделей составляют взаимно необходимые части современной структурной лингвистики.

ЛИНГВИСТИЧЕСКИЕ МОДЕЛИ

Глава I

ПОНЯТИЕ ЛИНГВИСТИЧЕСКОЙ МОДЕЛИ

В предыдущей части книги мы уже использовали, правда без каких-либо пояснений, термин «лингвистическая модель». Теперь мы должны будем рассмотреть это понятие гораздо более подробно, так как оно является центральным в современной структурной лингвистике, которую можно определить как науку о моделях языка по преимуществу.

Необходимость в моделировании возникает во всех тех научных областях, где объект науки недоступен непосредственному наблюдению. В таких случаях он обычно уподобляется некоему «черному ящику» <207>, <232>, о котором известно только, какие начальные материалы он получает «на входе» и какие конечные продукты он выдает «на выходе». Задача состоит в том, чтобы узнать содержимое «черного ящика» — тот скрытый от исследователя механизм, который осуществляет переработку исходных материалов в конечные продукты. Поскольку невозможно разобрать «черный ящик», не нарушив одновременно его функционирования, остается единственный путь к познанию нашего объекта: мы должны на основе сопоставления исходных и конечных данных построить образ объекта, т. е. выдвинуть гипотезу о его возможном устройстве и реализовать ее в виде логической машины, способной перерабатывать некоторый материал точно так же, как это делает сам «черный ящик». Если наше логическое устройство действительно функционирует аналогичным образом, оно является аппроксимацией, или моделью объекта, и мы можем считать, что заложенный в него механизм во всех существенных деталях совпадает с механизмом, содержащимся в «черном ящике». Мы добиваемся понимания механизма, недоступного

нашему наблюдению, уподобляя его механизму, устройство которого нам хорошо известно. Таким образом, смысл моделирования состоит в том, чтобы вместо скрытых от нас свойств объекта изучить заданные в явном виде свойства модели и распространить на объект все те законы, которые выведены для модели.

Первоначально проблема «черного ящика» возникла в электротехнике, но аналогичная ситуация имеет место во многих других областях, например в физиологии высшей нервной деятельности и атомной физике: мы не можем наблюдать непосредственно ни деятельности мозга, ни процессов, протекающих в атоме. В этом отношении положение лингвиста ничем не отличается от положения физиолога или физика: единственной реальностью, с которой лингвист непосредственно имеет дело, является текст, а интересующие его механизмы языка, лежащие в основе речевой деятельности человека, не даны ему в прямом наблюдении. Поэтому и в лингвистике одним из основных средств познания объекта является построение моделей.

Рассмотрим некоторые наиболее важные свойства моделей, в том числе лингвистических.

1. Моделировать можно только такие явления, существенные свойства которых исчерпываются их структурными (функциональными) характеристиками и никак не связаны с их физической природой. Круг таких явлений, по-видимому, гораздо более широк, чем думали раньше. Некоторые авторитетные исследователи относят к их числу даже процессы жизни и мышления <102>, <233>.

Если верно сосюрювское понимание языка, то к числу явлений, существенные признаки которых сводятся к их функциональным, или структурным, свойствам, относится и язык.

Моделью объекта, для которого существенными являются только его функциональные свойства, должно считаться любое устройство, функционально похожее на него. Иными словами, от модели требуется только, чтобы ее поведение было похоже на поведение объекта; материал, в котором она реализована, может отличаться (и практически почти всегда отличается) от материала, из которого построен объект. «Моделирование способа организации материальной системы,— пишет по этому поводу А. Н. Колмогоров,— не может заключаться ни

в чем ином, как в создании из других материальных элементов новой системы, обладающей в существенных чертах той же организацией, что и система моделируемая» <82>. С этой точки зрения совершенно не обязательно было бы требовать, чтобы, например, модель русского спряжения была непременно реализована в той же субстанции, что и сам объект (русское спряжение), т. е. закодирована состояниями нервных клеток мозга; она с равным успехом может быть записана карандашом на бумаге или набита на перфокарты и реализована в виде импульсов в электронной вычислительной машине. Во всех этих случаях содержащиеся в ней правила должны быть признаны правилами русского спряжения, если результаты ее работы во всех существенных деталях совпадают с соответствующими результатами деятельности человеческого мозга.

Функциональная точка зрения интересна потому, что делает задачу научного описания мира задачей конечной сложности. Выделение структурных свойств объекта в качестве наиболее существенных его свойств позволяет создать теорию данной структуры, равно применимую к объектам любой физической природы, если в их основе лежит та же структура. Таким образом, исследователь избавляется от необходимости придумывать новую теорию всякий раз, когда он сталкивается с реализацией той же структуры в новой субстанции. Установив, что в каких-то отношениях еще не изученный объект ведет себя так же, как объект, хорошо изученный и обеспеченный теорией, исследователь может попытаться распространить эту теорию (со всеми доказанными в ней теоремами) на первый объект, даже если субстанциально эти объекты совершенно непохожи друг на друга. В качестве классического примера обобщения такого рода можно привести теорию колебаний: колеблющиеся системы описываются одной и той же моделью (одними и теми же уравнениями) независимо от того, являются ли колебания акустическими, механическими или электромагнитными.

Влияние этих идей испытала в последнее время и лингвистика; в частности, Н. Хомский, основоположник современной структурной (математической) лингвистики, показал, что так называемые «порождающие грамматики», во всяком случае некоторые их типы, могут рассматри-

ваться как конечные автоматы, и поэтому к ним применима хорошо разработанная в математике теория конечных автоматов.

Все другие свойства модели связаны с тем ее основным свойством, что она является функциональной аппроксимацией объекта.

2. Модель всегда является некоторой *идеализацией* объекта. Реальные явления очень сложны. Чтобы понять эти сложные явления, необходимо начать с изучения самых простых и общих случаев, даже если они никогда не встречаются в чистом виде, и от них продвигаться к более сложным и специальным случаям. Этот прием не нов. Он знаком всем нам со школьных лет, когда мы решали задачи по физике и химии — не реальные задачи, которые ставит перед исследователем природа, а искусственные и упрощенные логические загадки, в которых описываются идеальные ситуации, практически не встречающиеся в чистом виде. Решая эти упрощенные задачи, в которых мы имели дело лишь со схемами вещей, мы тем не менее усваивали не научную фикцию, а важные теоретические истины.

На идеализации объекта основан существующий в криминалистике метод опознания преступника по его «словесному портрету» — описанию тех черт лица, которые связаны исключительно со строением черепа человека. В их число не входит цвет волос и кожи, цвет и выражение глаз и т. п.; оказывается, что ученый может ими пренебречь, хотя, например, для художника они представляют первостепенный интерес. Тот факт, что строение черепа и скелета определяет все существенные черты физического облика человека, был использован советским археологом и скульптором М. М. Герасимовым при разработке оригинального метода реконструкции внешности, с успехом применявшегося в археологии и криминалистике.

Чтобы понять механику движения, остававшуюся загадкой в течение тысячелетий, физики должны были предположить, что при отсутствии внешних влияний (трения, неровностей пути) становится возможным бесконечное движение тела по прямой линии, хотя практически такое движение никогда не может быть осуществлено, так как нельзя устранить все внешние влияния
<231, 42>.

Мы привели эти примеры несомненно удачной идеализации объекта, заимствованные из самых различных областей интеллектуальной деятельности человека, в надежде на то, что на их фоне нижеследующие лингвистические примеры будут рассмотрены читателем без предубеждения.

Если попросить кого-либо привести пример атрибутивной связи в русском языке, то почти всегда будет дано словосочетание типа *большой дом, красная роза, железная кровать*. Мало вероятно, чтобы в качестве первого примера атрибутивной связи привели словосочетания типа *прибавка четыре центнера, глубина 700 метров, стена длиной 600 метров, бег 500 метров* и т. п. (примеры заимствованы нами из <216>). Было бы, конечно, хорошо создать теорию, объясняющую все типы атрибутивной связи, в том числе и те весьма редкие, хотя и развивающиеся типы, которые в настоящее время находятся на границах литературной нормы (ср. последние четыре примера, где более нормальными были бы словосочетания *прибавка в четыре центнера, глубина в 700 метров, стена длиной в 600 метров, бег на 500 метров*). Однако, если нет теории, объясняющей самые обычные и массовые типы атрибутивной связи, представляется вполне оправданной такая идеализация объекта, которая исключает подобные раритеты из картины русской грамматики.

Идеализация такого рода неизбежно приводит к огрублению живого явления и схематизации фактов (ср. метод «словесного портрета»), которую сторонники классического языкознания склонны считать недопустимым насилием над языком. Но без такой схематизации научное описание невозможно; научная концепция того или иного явления «скорее диаграмма, чем картина» <304, 8>, и ученый, который стал бы настаивать на внесении в эту диаграмму всех фактов, касающихся данного явления, не смог бы справиться ни с одной научной задачей.

В рассмотренном нами случае некоторые реально существующие факты русской речи признаются «несуществующими». Гораздо чаще идеализация объекта сводится к тому, что признаются существующими некоторые реально не наблюдаемые факты. Мы рассмотрим ниже несколько примеров этого более классического типа идеализации.

Относительно давно известен в лингвистике принцип «подстановки», или «подразумевания», недостающих членов

предложения, который представляет собой простейший случай идеализации лингвистических объектов. Принцип подстановки часто используется в синтаксисе при описании некоторых типов предложений, на первый взгляд совершенно изолированных и занимающих особое место в синтаксической системе языка. В качестве примера можно упомянуть два основных подхода к объяснению так называемых безличных предложений русского языка. В большинстве грамматик они рассматриваются как особый класс предложений, утративших связь с личными предложениями. С другой стороны, предпринимались попытки, начиная с М. В. Ломоносова и А. Х. Востокова и кончая А. А. Шахматовым и В. А. Богородицким, истолковать их как результат эллипсиса, сфера действия которого в естественных языках, по-видимому, шире, чем принято было думать раньше. У В. А. Богородицкого, например, мы находим интересную мысль о том, что одночленные (безличные) предложения можно рассматривать как продукт сокращения нормальных двучленных предложений: *Морозит* ← *Мороз морозит* <20>. В современной лингвистике аналогичные соображения были с успехом использованы для принципиально нового описания соответствующего фрагмента русской грамматики (см. часть III).

Принцип подстановки является частным случаем более общего принципа конструирования фраз, не встречающихся в речи, но необходимых для объяснения всей совокупности языковых фактов. Так, для объяснения инфинитивных предложений с инфинитивом субъекта и объекта конструируются фразы типа *Я хочу, чтобы я пил* → *Я хочу пить*; *Я велю, чтобы он пришел* → *Я велю ему прийти*. Для объяснения сравнительных предложений типа *Он выше меня* рассматриваются предложения типа *Он выше, чем я высок* <195>. Для объяснения придаточных определительных предложений типа *Он встретил сына, которого не видел 20 лет* конструируются предложения вроде *Он встретил сына, которого сына он не видел 20 лет* и т. п. Все это имеет глубокий лингвистический смысл; идеальные фразы, которые реально в речи не встречаются, в принципе могли бы существовать и часто отражают либо прошлые, либо будущие этапы развития языка. Некоторые из них, например *Я хочу, чтобы я пил* или *Я велю, чтобы он пришел*, являются вполне нор-

мальными фразами, но только лишенными естественности, идиоматичности, которая в глазах многих является основным признаком доброкачественности языкового материала.

В связи с этим в правильном языковом материале стоит, по-видимому, различать то, что обычно говорят, и то, что может быть сказано, хотя обычно не говорится. По мнению С. Карцевского, трехчленная пассивная конструкция в русском языке является искусственным образованием <324>, так как в естественной речи почти никогда не встречается; тем не менее грамматисты самых различных направлений признают пассивные предложения с творительным падежом действующего лица нормальным языковым материалом. Известный английский грамматист Г. Суит рассматривал в своей «Новой английской грамматике, логической и исторической» <364> глагольные формы *I shall have been seeing, I shall have been being seen*, практически не встречающиеся в естественных условиях, но выводимые по грамматическим правилам¹. Любопытно, что, когда некий иностранец спросил у Г. Суита, можно ли сказать по-английски *an elegant supper* — «элегантный ужин», Г. Суит, будучи лингвистом, и притом последовательным, ответил, что английский язык — свободный, и сказать так можно, хотя он не припоминает случая, когда бы он сам назвал ужин элегантным.

Наиболее важным случаем идеализации лингвистического объекта является предположение, что число предложений в языке бесконечно и что длина предложения в принципе ничем не ограничена (т. е. возможны предложения, длина которых будет больше любой наперед заданной). Мы знаем, что ф а к т и ч е с к и число написанных или произнесенных предложений на любом языке конечно, хотя и очень велико; однако, чтобы объяснить способность говорящего строить совершенно новые, раньше никогда не произносившиеся и не писавшиеся предложения, мы должны рассмотреть не эту реально наблюдаемую ситуацию, а некоторую идеальную ситуацию, которая только и может дать нам ключ к решению задачи. Точно так же, хотя все реально произнесенные и написан-

¹ Ср. свыше полутора сотен форм английского глагола, выводимых в исчислении К. И. Бабицкого и Е. Л. Гинзбурга <15>; ср. <248>.

ные фразы на любом языке были фразами конечной длины, в строе естественных языков нет правил, ограничивающих их длину. Ввиду важности этого последнего пункта мы подробно рассмотрим относящиеся к делу примеры.

В одном фельетоне, посвященном стилю литературно-критических статей, высмеивался некий вымышленный автор, громоздящий один родительный падеж на другой: *Суть смысла темы раздела книги художника периода расцвета...* Стилистически это словосочетание никуда не годится, но грамматика русского языка допускает построение и более длинных именных словосочетаний с последовательным подчинением существительных в родительном падеже, а словосочетания, содержащие 4—5 существительных, встречаются настолько часто, что даже не вызывают удивления, ср. *Работники аппарата высшего органа законодательной власти республики...* («Известия», 3/VI-64); *Понимание необходимости сочетания уважения суверенитета ГДР...* («Правда», 3/V-62).

Аналогичным образом строятся словосочетания с последовательным подчинением инфинитивов, структурно родственные только что рассмотренным именным словосочетаниям, ср. *Соединенные Штаты должны перестать пытаться различать диктаторов и демократические режимы...* («Известия», 19/III-64). С. Карцевский <324> приводит еще более любопытный пример, звучащий достаточно непринужденно даже с наращением еще одного (последнего) члена: *Я не мог решиться поручить ему пойти просить вас пожаловать к нам отобедать.*

Последним весьма продуктивным типом подчинительных словосочетаний, которые могут практически бесконечно расти в длину, являются атрибутивные словосочетания вида *ее смеющийся старший брат, его старая железная двуспальная кровать*¹.

До сих пор мы имели дело со словосочетаниями. Однако предложение может расти в длину и за счет своих собственных ресурсов. Оставляя в стороне тривиальный случай сложносочиненных предложений, укажем, что почти все типы сложноподчиненных предложений допускают грамматически ничем не ограниченный рост; ср. ложно-многозначительные штампы современной прозы вроде следую-

¹ Одним из первых обратил внимание на словосочетания этого типа А. де Гроот <284, 10>.

щего: *Причем вру и чувствую, что она понимает, что я вру. И даже хуже: она понимает, что я понимаю, что она понимает, что я вру* (С. Г а н с о в с к и й).

Итак, если устранить действие таких подобных трению в мире физической действительности факторов, как ограниченный объем человеческой памяти, ограниченная продолжительность человеческой жизни и т. п., то следует признать, что грамматические правила языка допускают построение предложений сколь угодно большой длины. Поэтому более «проницательной» и глубокой будет та грамматика, которая ориентируется на идеализированную ситуацию, а не та, которая исходит из реально наблюдаемого факта конечной длины предложений. Такая идеализация многократно упрощает описание: совершенно очевидно, что нельзя сформулировать простых правил, ограничивающих длину предложения, и очень сомнительна возможность сформулировать сложные правила такого рода, которые были бы конечными.

3. Обычно модель оперирует не понятиями о реальных объектах, а к о н с т р у к т а м и, т. е. понятиями об идеальных объектах, не выводимыми непосредственно и однозначно из опытных данных, но построенными «свободно» на основании некоторых общих гипотез, подсказанных совокупностью наблюдений и исследовательской интуицией. Всякая модель является конструкцией, логически выведенной из гипотез с помощью определенного математического аппарата <207>.

Прежде чем иллюстрировать это положение языковым материалом, мы сошлемся на бесспорный пример плодотворности рассматриваемого здесь принципа конструирования теоретических понятий, заимствованный из области физики. Известно, что непосредственно наблюдаемое нами пространство трехмерно. Однако в созданной А. Эйнштейном теории относительности рассматривается не трехмерное, а четырехмерное пространство, причем в современных интерпретациях этой теории за четвертую координату принимается не время, а м н и м а я величина ($\sqrt{-c^2 t^2} = \sqrt{-1} \cdot ct$, где c — скорость света, а t — время в секундах). В то время как измеренная в трехмерном пространстве длина отрезка получает различные значения в зависимости от скорости наблюдателя, длина отрезка, измеренная в четырехмерном пространстве, инвариантна (неизменна) относительно системы координат.

«Многие законы приобретают в такой системе чрезвычайно простой вид, удобный для любых построений» <77, 32>. Четвертое измерение теории относительности — это «удобный и наглядный прием изображения законов природы, позволяющий математически сформулировать связь, существующую между пространством и временем» <77, 32>. Имея в виду подобные построения, А. Эйнштейн писал: «Я убежден, что чисто математические конструкции позволяют найти понятия и связывающие их законы, которые дают ключ к явлениям природы. Опыт, разумеется, может руководить нашим выбором нужных математических понятий, но он практически не может быть источником, из которого они вытекают. В известном смысле я считаю истиной, что чистая мысль способна ухватить реальное, как об этом мечтали древние» <83, 78>; «...именно в математике,— говорит он в другом месте,— содержится действительно творческий принцип» <83, 105>.

Типичным конструктом в лингвистике является понятие нулевой флексии (или нулевого варианта флексии) для слов типа *депо*, *шоссе*; такие «нули» непосредственно не наблюдаются, но свободно конструируются исследователем для объяснения наблюдаемых фактов, в частности факта согласования этих слов с прилагательным в роде, числе и падеже. Укажем на еще один конструкт, издавна используемый в самых традиционных описаниях русского языка. Речь идет о грамматических категориях прилагательных и других согласуемых слов. Мы воспользуемся очень ясным изложением этого вопроса, принадлежащим А. А. Зализняку: «Необходимо учитывать, что признание у прилагательных и других согласуемых слов большего или меньшего числа категорий, по которым происходит согласование, не вытекает однозначно из фактов, а отражает определенные принципы описания. В самом деле, у русских прилагательных различаются не более 13 внешне различных словоформ (имеются в виду полные формы положительной степени), например: *новый*, *-ая*, *-ое*, *-ого*, *-ой*, *-ому*, *-ую*, *-ым*, *-ою*, *-ом*, *-ие*, *-ых*, *-ыми*. Известно, что в этих словоформах нельзя найти отдельного внешнего выражения для числа, или падежа, или рода, а все эти значения выражены синкретически. Поэтому вполне естественно следующее описание. Существует всего одна грамматическая категория прилагательных, по которой происходит согласование. Она при-

нимает 13 различных значений, т. е. прилагательные имеют всего 13 грамматических форм. Правила выбора форм при атрибутивной связи задаются в виде таблицы, где каждому набору грамматических значений существительного соответствует номер формы прилагательного, например: в. ед. муж. одуш.—№ 4 (-ого). Такое описание дает максимально простую парадигму прилагательного, но сложные правила выбора формы подчиненного слова. Общепринятым является другое описание, а именно: прилагательному приписывают все те грамматические категории существительного, которые хотя бы в одном случае влияют на выбор формы прилагательного. 13 форм разносятся по 36 (а после признания категории одушевленности — по 72) клеткам. Получается более сложная парадигма, с многочисленными случаями омонимии, но зато достигается максимально простое правило выбора формы при атрибутивной связи: «грамматические значения прилагательного повторяют грамматические значения подчиняющего существительного» (откуда и само слово «согласование»). Как указал (устно) И. А. Мельчук, выбор второго решения в значительной мере определен также тем, что при субстантивации грамматические характеристики, приписываемые прилагательному в соответствии с этим решением, автоматически превращаются в одноименные характеристики существительного» <62, 29>.

Следовательно, и в этой области современная структурная лингвистика развивает, весьма последовательно и решительно, принципы, не чуждые классическому языкознанию.

В идеале лингвистические конструкты суть понятия, построенные без непосредственного обращения к субстанции, фонетической или семантической, тех явлений, для изучения которых они созданы. Из этого, однако, не следует, что значение не может быть объектом лингвистической теории; поскольку основная функция языка состоит в том, чтобы передавать смысл, построение моделей семантики является «неотъемлемой частью полного описания языка» <263, 284>. «Описать язык, не установив на каком-то этапе его значений,— говорил известный американский психолог и психолингвист Дж. Кэрролл,— это то же самое, что разработать код без ключа к нему» <259, 19> (см. также главу 2).

4. Всякая модель, в том числе лингвистическая, должна быть формальной. Модель считается формаль-

ной, если в ней в явном виде и однозначно заданы исходные объекты, связывающие их утверждения и правила обращения с ними (правила образования или выделения новых объектов и утверждений). В идеале всякая формальная модель является математической системой. Поэтому в некотором смысле понятие формальности равнозначно понятию математичности, точности, или однозначности ¹.

Формальность, точность, однозначность — это свойство языка, на котором излагается теория. Само по себе это свойство не обеспечивает совпадения предсказаний формальной теории с объективными экспериментальными данными. Точность теории делает возможной постановку недвусмысленных экспериментов, которые способны подтвердить или опровергнуть ее, но никакой необходимой логической связи между точностью и истинностью теории нет и не может быть.

Формальная модель связывается с опытными данными посредством той или иной интерпретации. Дать интерпретацию модели — значит указать правила, вероятностные или строгие, подстановки объектов некоторой предметной области, например языка, вместо объектов (символов) модели ².

Из того, что было сказано выше о свойствах модели как функциональной аппроксимации объекта, следует, что число возможных интерпретаций данной модели в принципе ничем не ограничено и, во всяком случае, больше одного. Пусть, например, в модели рассматриваются элементы $a_1, a_2, a_3, \dots, a_n$ и $b_1, b_2, b_3, \dots, b_m$ и цепочки вида

$a_j b_i, a_i b_j a_j, b_i a_j$ и т. п. (стрелка показывает, что b_i

¹ Формализация некоторой области содержательных представлений сопровождается обычно введением символических обозначений, но, конечно, как следует из сказанного выше, не сводится к нему. В этой связи следует подчеркнуть, что распространившееся в последнее время увлечение символами, за которыми не стоят точно (однозначно) определенные понятия, не имеет ничего общего с научным методом. Научообразная символика не может заменить работу мысли и способна только дискредитировать большое и серьезное дело в глазах людей, искренне им интересующихся, но мало осведомленных.

² Логически интерпретация является совершенно особым этапом исследования, хотя практически она не всегда излагается отдельно от самой формальной модели. Об этом следует помнить при знакомстве с материалом частей III и IV.

является главным, а a_i и a_j — зависимым элементом); модель может быть интерпретирована фонологически: вместо a_1, a_2, \dots, a_n мы подставляем согласные, вместо b_1, b_2, \dots, b_m — гласные и интерпретируем цепочки указанного вида как фонологические слоги с гласным в вершине слога. Модели может быть дана и грамматическая интерпретация. Тогда вместо a_1, a_2, \dots, a_n мы подставляем группу существительного, вместо b_1, b_2, \dots, b_m — группу глагола в личной форме и интерпретируем цепочки вида $a_j b_i, a_i b_j a_j$ и т. п. как предложения со сказуемым в качестве его вершины (ядерного элемента) <90>, <152>.

В рассмотренном здесь примере модель интерпретировалась для различных предметных областей (фонологии и грамматики) внутри одного языка. Абстрактная модель может быть интерпретирована и на однотипном материале (например, грамматическом) разных языков. Представим, например, модель, в которой рассматриваются, в частности, элементарные символы N, V, A, D и производные символы $N(N), N(V), N(A), N(D), V(N), V(V), V(A), V(D)$ и т. д. Интерпретируем N как основу непроизводного существительного, V — как основу непроизводного глагола, A — как основу непроизводного прилагательного, D — как основу непроизводного наречия. Тогда $N(V)$ интерпретируется как основа инфинитива и отглагольного существительного на материале русского языка и как основа инфинитива, отглагольного существительного и герундия на материале английского <211>.

Эти положения дают нам ответ на вопрос о том, каким образом свободные, или идеальные, математические конструкции могут объяснять поведение некоторых объектов вполне определенной природы. Модель, построенная для объяснения некоторого эмпирического материала, но не допускающая ни одной строгой интерпретации, является научной фикцией; она должна быть отброшена и заменена новой моделью. Модель тем эффективнее, чем шире ее предметная область, т. е. чем больше число допускаемых ею интерпретаций ¹.

¹ Впрочем, следует иметь в виду, что существует обратное отношение между широтой модели (числом допускаемых ею интерпретаций) и ее богатством: чем шире предметная область модели, тем беднее ее содержание (тем меньше ее моделирующая сила), и наоборот.

5. Всякая интерпретированная модель, в том числе лингвистическая, должна обладать свойством *э к с п л а н а т о р н о с т и*, или объяснительной силы. Считается, что модель обладает этим свойством, если она 1) объясняет факты или данные специально поставленных экспериментов, которые необъяснимы с точки зрения старой теории, 2) предсказывает неизвестное раньше, но принципиально возможное поведение объекта, которое позднее подтверждается данными наблюдения или новых экспериментов. И в том и в другом случае объяснительная сила модели тем больше, чем полнее мера совпадения предсказаний с экспериментальными данными.

Классической иллюстрацией первого случая является специальная (частная) теория относительности А. Эйнштейна, объяснившая знаменитый опыт А. Майкельсона, результаты которого казались совершенно загадочными с точки зрения доэйнштейновой физики. Классическими иллюстрациями второго случая являются принадлежащая ему же общая теория относительности, основной вывод которой был подтвержден два года спустя после ее формулировки экспериментом А. Эддингтона; открытие Д. И. Менделеева, предсказавшего существование ряда в его время неизвестных элементов; теоретические расчеты У. Леверрье, из которых с неизбежностью вытекал вывод о существовании в солнечной системе еще одной планеты (Нептуна), позднее действительно открытой исследователями.

В лингвистике также есть прецеденты обоего рода. Ввиду исключительной важности рассматриваемого здесь вопроса мы остановимся на них несколько более подробно.

В 1946 году была опубликована широко известная статья Г. О. Винокура о русском словообразовании <37>, в которой излагались, в частности, разработанные им принципы морфологического анализа производных слов. Для нас представляет интерес только анализ слов с уникальными основами типа *малина*, *смородина*, *буженина*, *аптека* и т. п. и слов с уникальными суффиксами типа *пас-тух*, *жен-их*, *рис-унок*, *корол-ева*, *поп-адья*, *пе-сня*, *враж-да* и т. п. Слова первого типа он считал непроеизводными, несмотря на наличие семантически близких к ним слов с ясно выраженной делимостью на основу и аффикс, ср. *буженина* — *баран-ина*, *свин-ина*, *телят-ина*; *аптека* — *фото-тека*, *карто-тека*. Слова второго типа он считал

производными. «Неравноправие» уникальных основ и уникальных суффиксов Г. О. Винокур объяснял различием в значениях корневых и аффиксальных морфем. Значение аффикса является чисто дифференциальным; оно устанавливается в противопоставлениях типа *пас-ти* — *пас*, *нес-ти* — *нес*, *пас-ти* — *пас-тух* и т. п. В отличие от этого, для установления значения основы необходимо соотнести ее не с другой основой, а с чем-то вне языка; ее значение является не дифференциальным, а вещественным. Поскольку мы не знаем, какой предмет реального мира обозначается элементом *бужен-*, или *мал-*, или *смород-*, мы не можем выделить его в качестве первичной (непроизводной) основы слов *буженина*, *малина*, *смородина*. Чтобы такое разложение было возможно, необходимо существование хотя бы еще одного слова, где встречается данная первичная основа.

Два года спустя появилась полемическая статья А. И. Смирницкого <165>, в которой излагались принципы морфологического анализа, реализующие другую содержательную гипотезу о структуре производных слов. А. И. Смирницкий исходил из того, что и у корней, и у аффиксов имеется дифференциальное значение (это не мешает ни тем, ни другим иметь и некоторое вещественное значение). Поэтому корень и аффикс в составе основы рассматривались им на равных основаниях. Общие условия разложимости основы были сформулированы А. И. Смирницким следующим образом: пусть дана основа *L* и внутри нее — звуковые отрезки *A* и *B*. Эти отрезки имеют значение (и, следовательно, основа *L* разложима), если 1) хотя бы один из них встречается не только в *L*, но и в какой-либо другой основе *M* с другим звуковым отрезком *C* или с нулем; 2) *L* и *M* относятся к таким предметам и явлениям, которые обладают отчетливо выделяемыми общими признаками (общим значением): «... В результате осознания этих признаков образуется значение этого общего звукового отрезка, тогда как звуковые отрезки, — в том числе и нуль, — дифференцирующие основы *L* и *M*, неизбежно получают те значения, которыми осмысляются различительные признаки предметов, явлений и т. п., обозначаемых посредством *L* и *M*» <165, 25>. Если в *смород-ин-а* и *мал-ин-а -ин-* значит «ягода», то *мал-* и *смород-* значат «то в малине и смородине, что отличает их друг от друга и от прочих ягод» <165,

24>¹. Аналогичный, а иногда и буквально совпадающий анализ слов с уникальными основами мы находим в работах таких представителей структурной лингвистики, как Л. Блумфильд, Г. Глисон, Дж. Гринберг, З. Харрис, Ч. Хоккет и др.².

Итак, перед нами две разные миниатюрные теории, объясняющие некоторый кусочек языковой действительности, и на первый взгляд неясно, которую из них следует предпочесть. Мы должны были бы узаконить обе эти теории в качестве столь привычных для лингвиста «различных точек зрения по данному вопросу», если бы не существовало эмпирического материала, который одна из них объясняет, а другая нет. Теория Г. О. Винокура не объясняет широко известных по материалам самых различных языков фактов так называемого обратного словообразования. Существо процессов обратного словообразования можно пояснить с помощью уже приводившегося примера выделения слова *зонт* из слова *зонтик*. Это последнее было заимствовано из голландского языка (ср. *zonnedek*) и в русском языке было сначала простым по структуре. Позднее оно ассоциировалось с уменьшительными типа *столик*, *мячик*, *шарик* и т. п. и само стало восприниматься как уменьшительное. В результате оно разложилось на уникальную основу *зонт-* и уменьшительный суффикс *-ик*, причем тот факт, что основа со временем вошла в состав непроизводного слова *зонт*, свидетельствует об осознании самостоятельности ее значения («то в *зонтике*, что отличает его от *столика*, *мячика*, *шарика* и т. п.»). Аналогичным образом выделились в английском языке слова *to beg* («просить») из *beggar* («нищий») [или *beguine* (член нищенствующего ордена)] и *to chauffe* («водить машину», «быть шофером») из *chauffeur* («шофер»). И в том и в другом случае основы являются уникальными, так как существительные *beggar*, *chauffeur* были заимствованы из французского языка (ср. французские *bag-*

¹ Слова данного типа анализируются и другими способами, отличными от способа Г. О. Винокура и способа А. И. Смирницкого; см., например, <156, 34> и <67, 88>.

² Сопоставляя слово *cran-berry* («клюква»), содержащее уникальную основу *cran-*, со словами *berry* («ягода»), *blackerry* («черная смородина») и другими подобными, Ч. Хоккет говорит: «...Значение *cran-*, таким образом, — это все то, что отличает клюкву от других видов ягод. Его, может быть, трудно описать, но легко продемонстрировать на фруктовом рынке» <309, 126>.

(h)ard, chauffeur) и не имели на английской почве никаких параллелей. Все эти факты непринужденно объясняются теорией А. И. Смирницкого и тех исследователей, которые придерживались аналогичных принципов морфологического анализа.

Перейдем к прецеденту второго рода — лингвистической теории, которая не только объясняет весь наличный материал, но и предсказывает до того не наблюдавшиеся факты. В качестве примера такой теории мы рассмотрим в несколько упрощенном виде соссюрговскую концепцию ларингальных, уже упоминавшуюся нами на стр. 3—4¹. К этой концепции Ф. де Соссюр пришел в 1879 году, изучая индоевропейские чередования гласных. Эти чередования можно иллюстрировать следующими греческими примерами:

Н а с т о я щ е е	П е р ф е к т	А о р и с т
<i>e</i>	<i>o</i>	<i>ø</i>
<i>pétomai</i> — „я лечу“	<i>perótēmai</i>	<i>eptómēn</i>
<i>ei</i>	<i>oi</i>	<i>i</i>
<i>peithō</i> — „я убеждаю“	<i>pēpoitha</i>	<i>ēpithon</i>
<i>er</i>	<i>or</i>	<i>r</i>
<i>dérkomai</i> — „я гляжу“	<i>dédorka</i>	<i>édracon</i>

Во всех случаях имеет место индоевропейское чередование *e—o—ø*, причем гласный корня может быть осложнен звуками *i, u, l, m, n, r*. Чередование идет по двум признакам: 1) передние — задние гласные (*e—o*) (так называемое качественное чередование); 2) гласный полного образования — нуль звука (так называемое количественное чередование). Эта великолепная картина нарушается из-за следующих, правда, менее частотных чередований, где появляются долгие гласные и исчезает чередование «гласный — нуль звука»:

<i>a</i>	<i>ō</i>	<i>a</i>
<i>phāmī</i> — „я говорю“	<i>phōnē</i> — „голос“	<i>phatōs</i> — „сказанный“
<i>histāmī</i> — „я кладу“		<i>statōs</i> — „положенный“
ср. лат. <i>dare</i> — „давать“	<i>dōnum</i> — „дар“	<i>datus</i> — „данный“
<i>stare</i> — „стоять“		<i>status</i> — „стояние“

¹ Она изложена Ф. де Соссюром в <353>. См. также ясное и простое изложение вопроса в <51> и <374>. В современной лингвистике основные выводы ларингальной теории уже не оспариваются, хотя некоторые специалисты не разделяют ряда ее положений <380> или относятся к ним с осторожностью <171>, особенно в связи с материалами хеттского языка, положение в котором оказалось сложнее, чем думали вначале.

Для объяснения этой нерегулярности выдвигались различные более или менее правдоподобные принципы, признание которых, однако, сильно усложняло картину. Ф. де Соссюр принял в качестве исходного наиболее обычный тип чередования *e—o—φ* (где за гласным корня могут следовать указанные выше звуки). Он допустил, что в серии *ā—ō—a* долгие гласные звуки образовались в результате стяжения краткого гласного корня с каким-то неизвестным звуком, не давшим никаких рефлексов в исторически засвидетельствованных (к тому времени) языках. Этот звук, который мы условно обозначим буквой *A*, имел, по Ф. де Соссюру, качество ларингального сонанта (он называл его сонантическим коэффициентом), и его фонетический эффект состоял в том, что он обращал гласный корня в гласный заднего ряда типа [a] или [o]. Если в безударном положении гласный корня исчезал (нулевая ступень чередования), то ларингальный, являющийся, по предположению, сонантом, отражался либо в виде *a*, либо *i*. В сильно упрощенном виде этот процесс можно представить себе следующим образом <374>:

<i>pheAmi</i> → <i>phāmī</i>	<i>phAtos</i> → <i>phatōs</i>
<i>histeAmi</i> → <i>histāmī</i>	<i>stAtos</i> → <i>statōs</i>
<i>phoAne</i> → <i>phōnē</i>	<i>stAtus</i> → <i>status</i>
<i>doAnum</i> → <i>dōnum</i>	<i>dAtus</i> → <i>datus</i> ¹

«Преимущество такого анализа перед классическим,— пишет по этому поводу Л. Ельмслев,— состояло, во-первых, в том, что он давал более простое решение проблемы, устраняя так называемые долгие гласные из системы, а с другой стороны, в том, что получалась полная аналогия с чередованиями гласных, которые до тех пор рассматривались как нечто фундаментально отличное... Этот анализ был произведен исключительно по внутренним причинам, с целью проникнуть глубже в основную систему языка; он не был основан на каких-нибудь очевидных данных самих сравниваемых языков; он был внутренней

¹ Более точно <353, 135>, Ф. де Соссюр считал, что всякий индоевропейский корень имел только один гласный *a*₁ (соответствующий греческому *e*), который в некоторых условиях переходил в *a*₂ (соответствующий греческому *o*). Ф. де Соссюр постулировал два сонантических коэффициента, *A* и *Q*, причем качество получавшегося в результате стяжения долгого гласного (*ā* или *ō*) зависело не от качества корневой гласной, а от качества ларингального.

операцией, произведенной в индоевропейской системе» <51, 51>.

Ф. де Соссюру был 21 год, когда он написал свою работу¹. Она опередила современную ему науку не менее чем на 50 лет. Лишь в 1927 году, после дешифровки хеттского языка, были обнаружены первые факты, предсказанные теорией Ф. де Соссюра. Е. Курилович указал, что хеттское *ḥ* является рефлексом индоевропейского ларингального сонанта, и привел следующие параллели, подтверждавшие правильность теории: лат. *pāscunt* — «они защищают», зап.-тохарское *pāskem* — «они защищают», хет. *pahsanzi* — «они защищают»; в хеттском — краткий гласный с последующим ларингальным; лат. *novāre* — «делать новым», хет. *newahh* — «делать новым»; греч. *laos* — «армия», хет. *lahha* — «война».

Много примеров такого рода знает история лингвистической дешифровки. Ж. Ф. Шампольон дешифровал египетское иероглифическое письмо в 1824 году, но правильность предложенной им «модели» была бесповоротно подтверждена только в 1866 году, когда его последователь Р. Лепсиус нашел в Египте, в местечке Сан, камень с текстом на трех языках («Канопский декрет») и, переводя египетский текст на греческий язык по методу своего учителя, получил перевод, полностью совпавший с греческим текстом декрета. Подтверждение правильности дешифровки хеттского иероглифического письма, начатой Б. Грозным, было дано более 20 лет спустя Х. Т. Боссертом, которому посчастливилось найти в 1947 году билингву с надписями, сделанными хеттским иероглифическим и финикийским письмом. Дешифровка линейного письма В, предложенная М. Вентрисом, была подтверждена несколько лет спустя У. Блегеном, который прочел, пользуясь методом М. Вентриса, табличку из Пилоса <46>.

¹ Чтобы читатель мог составить некоторое представление о величии этого открытия, мы рекомендуем ему попытаться решить последнюю задачу из уже упоминавшейся статьи А. А. Зализняка <61, 147>. Словоформы подобраны в ней таким образом, что «становится возможным строго сформулировать проблему, впервые поставленную и разрешенную де Соссюром... Таким образом, при решении этой задачи читатель должен самостоятельно повторить открытие де Соссюра». Наше изложение облегчит, но не обесмыслит ее.

Во всех этих случаях имела место реконструкция исторических фактов, причем ее правильность подтверждалась материалом, который в силу тех или иных причин, чаще всего случайных, обнаруживался значительно позднее. Эта редкостно благоприятная экспериментальная ситуация, но без присущего ей элемента случайности легко проецируется и на синхронические исследования. С указанной точки зрения между диахроническими и синхроническими исследованиями есть лишь одно принципиальное различие: в диахронических исследованиях реконструируются уже бывшие факты, а в синхронических исследованиях конструируются существующие и возможные будущие факты.

Это создает широкие возможности для экспериментов, на которые одним из первых обратил внимание Л. В. Щерба <224>, <225>. Он считал, что задача лингвиста не исчерпывается составлением грамматики и словаря на основании ограниченного материала: «Построив из фактов этого материала некую отвлеченную систему, необходимо проверять ее на новых фактах, т. е. смотреть, отвечают ли выводимые из нее факты действительности. Таким образом, в языкознание вводится принцип эксперимента. Сделав какое-либо предположение о смысле того или иного слова, той или иной формы, о том или ином правиле словообразования или формообразования и т. п., следует пробовать, можно ли сказать ряд разнообразных фраз (который можно бесконечно множить), применяя это правило. Утвердительный результат подтверждает правильность постулата... Но особенно поучительны бывают отрицательные результаты: они указывают или на неверность постулированного правила, или на необходимость каких-то его ограничений, или на то, что правила уже больше нет, а есть только факты словаря, и т. п. ...В возможности применения эксперимента и кроется громадное преимущество с теоретической точки зрения изучения живых языков. Только с его помощью мы можем действительно надеяться подойти в будущем к созданию вполне адекватных действительности грамматики и словаря» <225, 308—309> ¹.

¹ Большое значение «отрицательному языковому материалу» придавали В. А. Богородицкий <20>, А. Фрей <273>, Ш. Балли <17>, видевшие в патологии преувеличение нормы или зародыш будущего развития.

На той же идее основан предложенный Н. Хомским способ экспериментальной проверки объяснительной силы лингвистической модели: модель должна уметь строить не только те языковые объекты, которые уже встречались в речевой практике говорящих, но и объекты, принципиально допустимые, хотя и не встречавшиеся еще в речевой практике. Аналогичным образом модель, имитирующая речевую деятельность слушающего, должна обладать способностью анализировать не только те речевые произведения, которые послужили в качестве исходного материала при ее разработке, но и другие правильные речевые произведения. Только такие модели могут объяснить способность говорящего строить любые новые предложения (за исключением неправильных) и способность слушающего понимать любые новые предложения (опять-таки за исключением неправильных). Отметим, что такие модели способны объяснить и процесс усвоения языка ребенком <201>.

Эти указания, конечно, не решают исключительно сложного и интересного вопроса об экспериментальных способах проверки предсказаний модели и определения ее объяснительной силы. В идеале мы имеем право утверждать, что предсказания модели подтверждаются опытными данными, только в том случае, если эти предсказания сформулированы вполне точно не только с качественной, но и с количественной стороны. История науки знает немало ложных теорий, несоответствие которых действительности выяснялось не потому, что они не предсказывали общую форму поведения объекта, а потому, что они неверно предсказывали его количество. К сожалению, в лингвистике только сейчас начинается разработка экспериментов, способных подтвердить не только качественные, но и количественные аспекты предсказаний¹.

Итак, построение модели предполагает:

- 1) фиксирование фактов, требующих объяснения,
- 2) выдвижение гипотез для объяснения фактов,

¹ Во всех этих случаях имелись в виду реш а ю щ и е эксперименты, подтверждающие теорию или опровергающие ее. Их не следует путать с теми экспериментальными приемами обработки сырого материала, которые издавна известны в языкознании (ср., например, систему, разработанную американскими дескриптивистами; см. также часть III, гл. 2).

3) реализацию гипотез в виде моделей, не только объясняющих исходные факты, но и предсказывающих новые, еще не наблюдавшиеся факты,

4) экспериментальную проверку модели <206, 108>.

Все начинается с опыта, и все кончается им, и если между предсказаниями и фактическим положением вещей имеется расхождение, то, в зависимости от того, насколько оно велико, модель уточняется, перестраивается или отвергается вовсе, и тогда надо все начинать сначала. Если же некоторая система положений вообще не допускает никакой экспериментальной проверки и не может быть доказана другими (например, дедуктивными) способами, она не имеет права называться ни моделью, ни теорией.

Глава 2

ТИПЫ ЛИНГВИСТИЧЕСКИХ МОДЕЛЕЙ

Типология лингвистических моделей находится в настоящее время в предварительной стадии разработки, и поэтому мы не сумеем дать исчерпывающей классификации всех возможных типов моделей¹. Свою задачу мы видим в том, чтобы указать основные признаки, по которым классифицируются модели, и кратко охарактеризовать особенности тех или иных классов.

Начнем с описания моделей, отличающихся друг от друга по характеру рассматриваемого в них объекта. С этой точки зрения можно различать три типа моделей <190>:

1) Модели, в которых в качестве объекта выступают конкретные языковые процессы и явления. Эти модели имитируют речевую деятельность человека. Первый серьезный шаг в их разработке был сделан пражскими структуралистами.

2) Модели, в которых в качестве объекта рассматриваются процедуры, ведущие ученого к обнаружению того или иного языкового явления. Эти модели в некоторой мере имитируют исследовательскую деятельность лингвиста; они именуются ниже моде-

¹ О типах лингвистических моделей см. <5>, <12>, <152>, <190>, <199>, <202>, <209>, <240>, <272>, <276>, <299>, <307>, <318>, <355>, <359> и др.

л я м и и с с л е д о в а н и я <118>. Первый серьезный шаг в их разработке был сделан американскими дескриптивистами, пытавшимися создать универсальные алгоритмы открытия грамматик естественных языков.

3) Модели, в которых в качестве объекта рассматриваются уже готовые лингвистические о п и с а н и я, а не речевая деятельность человека или исследовательская деятельность лингвиста. Если обычная грамматика является теорией конкретных языковых процессов, то модель третьего типа является теорией теории, или м е т а т е о р и е й. Первый серьезный шаг в разработке моделей третьего типа был сделан глоссематиками.

Модели, имитирующие речевую деятельность человека, являются наиболее важным типом собственно лингвистических моделей. По отношению к ним модели второго и третьего типов выполняют вспомогательную роль.

Модели, имитирующие исследовательскую деятельность лингвиста, предназначены для того, чтобы объективно обосновать выбор тех понятий, которыми он пользуется при изложении модели первого типа, например грамматики того или иного языка. В идеале, который пока никем не был достигнут, они сводят до минимума роль субъективного фактора в исследовании. Таким образом, они в некотором смысле являются мерилем правильности (истинности) моделей первого типа.

Эффективность модели первого типа можно повысить еще одним способом — не давая объективное обоснование используемым в ней понятиям, а сравнивая ее по некоторым критериям, в том числе экспериментальным, с другими аналогичными моделями. В этом и состоит назначение моделей третьего типа: они обеспечивают систему критериев и теоретических доказательств (метаязык), с помощью которых из нескольких предъявленных моделей, моделирующих одно и то же явление, мы можем выбрать лучшую.

Поскольку исследовательские модели в идеале предшествуют моделям конкретных языковых процессов, а метамоделю следуют за ними, мы будем в этой главе и в последующих частях книги держаться указанного в данном абзаце порядка изложения материала: мы начнем с рассмотрения исследовательских моделей, затем перей-

дем к моделям конкретных языковых процессов (моделям речевой деятельности человека) и расскажем в заключение об основных положениях метатеории.

Мы уже говорили о том, что модели второго типа имитируют исследовательскую деятельность лингвиста. Эта деятельность состоит в том, что из суммы наблюдений над речевыми произведениями (текстами) извлекается некоторое представление о способе их организации, т. е. с и с т е м е, порождающей их. Следовательно, имитировать деятельность лингвиста — значит обеспечить переход от совокупности текстов к лежащей в их основе системе. Система считается в достаточной мере изученной, если мы знаем: 1) ее элементарные единицы, 2) классы элементарных единиц, 3) законы сочетания элементов различных классов на всех уровнях анализа, включая семантический (ср. постановку этого вопроса у американских дескриптивистов, стр. 46—47).

Исследовательские модели можно подразделить на три класса в зависимости от того, какая информация используется в них в качестве исходной. В моделях первого класса в качестве исходной информации используется только т е к с т, и все сведения о системе, т. е. языке, порождающем этот текст, извлекаются исключительно из текстовых данных. Это классические дешифровочные модели. В моделях второго класса считается заданным не только текст, но и м н о ж е с т в о п р а в и л ь н ы х ф р а з данного языка. Практически это значит, что при разработке модели лингвист прибегает к помощи информанта, который по поводу каждой предъявляемой ему фразы должен говорить, правильна она или нет. Информантом может быть и сам лингвист, если он в совершенстве владеет изучаемым языком. Наконец, в моделях третьего класса считаются заданными не только текст и множество правильных фраз, но и м н о ж е с т в о с е м а н т и ч е с к и х и н в а р и а н т о в. Практически это значит, что информант должен определять не только правильность каждой предъявляемой ему фразы, но и о любых двух фразах говорить, значат ли они одно и то же или нет. Модели этого класса близки традиционным описаниям, и в дальнейшем мы на них не останавливаемся.

Перейдем теперь к моделям конкретных языковых процессов и явлений, которые иначе можно назвать м о д е л я м и р е ч е в о й д е я т е л ь н о с т и ч е л о

века или моделями владения языком. Прежде чем обсуждать типы таких моделей, следует уточнить, какой смысл вкладывается в слова «владение языком»; в частности, следует определить свое отношение к самому важному из всех лингвистических вопросов — вопросу о том, предполагает ли владение языком знание смысла (иными словами, входит ли значение в структуру языка или нет). Несколько замечаний по этому поводу было сделано раньше, на стр. 88, но вопрос настолько серьезен, что заслуживает несколько более детального обсуждения.

В первой части мы уже приводили мнение Л. Блумфильда о том, что конкретное значение слова не должно интересоваться лингвиста, так как оно не входит в структуру языка. Лингвист не может и не должен анализировать различие между *фламинго* и *колибри* (это — дело зоолога), между *метаном* и *этаном* (это — дело химика), между *нейтрино* и *нейтроном* (это — дело физика). Если бы мы держались иного мнения на этот счет, нам пришлось бы внести в толковые словари культурных языков около 300 000 химических терминов, около 600 000 энтомологических терминов и т. п. Тогда лингвистика должна была бы слиться с энциклопедическим сводом всех накопленных человечеством знаний. Между тем известно, что человек может в совершенстве владеть языком, либо вовсе не зная большей части научных и технических терминов и терминообразных слов, либо зная их чисто номинально и имея лишь смутное представление о том, что они значат. Ж. Вандриес заметил по этому поводу, что каждый культурный человек знает номинально несколько специальных словарей, иногда даже на нескольких языках, не зная сколько-нибудь точно значений соответствующих слов <29>. Из этого следует, что владение языком не предполагает знания всех семантических различий между всеми словами, и в какой-то мере Л. Блумфильд был прав, когда он утверждал, что значение не может быть проанализировано в рамках нашей науки.

Некоторые лингвисты, в особенности крайние представители американского дескриптивизма, сделали из этого вывод, что значение ни в каком смысле не входит в структуру языка и что владение языком, как таковым, проявляется у говорящего в способности строить грамматически правильные фразы, а у слушающего —

в способности понимать такие фразы. При этом под грамматической правильностью понимается соблюдение чисто формальных запретов: фраза *Она пришел* неправильна, а фраза *Квадрат выпил гипотенузу* совершенно безупречна. Для обоснования этого взгляда можно как будто сослаться на следующие довольно любопытные факты:

1) Легко определить, какого рода текст — художественный или научный — закодирован фразой *Под стурической стурой стурой стур стурается стуренность стураций о стурах между стурами и между стурыми стурениями этой стуры* (опыт Г. А. Лесскиса, проведенный по рекомендации В. А. Ицковича), хотя все корневые морфемы были заменены одной безразличной корневой морфемой, не имеющей значения в русском языке.

2) Каждому англичанину «понятны» стихи Льюиса Кэррола из его широко известной книги «Алиса в стране чудес»:

*'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.*

Хотя в этом четверостишии нет ни одного английского корня, сказочный персонаж Хампти-Дампти, слушавший его вместе с Алисой, объяснил, что *outgrabe* — прошедшее время глагола *to outgribe*, а Алиса сказала, что стихи наполняют ее голову мыслями, хотя она не знает в точности, какими именно.

3) Каждому человеку, владеющему русским языком, в каком-то смысле «понятна» придуманная Л. В. Щербой фраза, не содержащая ни одной лексической морфемы русского языка: *гло́бая кúздра штéко будланúла бокрá и кудря́чит бокрénка*¹; это понимание проявляется, в частности, в том, что он сумеет поставить вопросы ко всем членам предложения и ответить на них.

4) Каждый англичанин «поймет» придуманные Ч. Фризом предложения <274, 71>:

1. *Woggles ugged diggles.*
2. *Uggs woggled diggs.*
3. *Woggs diggled uggles.*

¹ Цитируется по <187>.

и сумеет вывести по правилам английской грамматики другие фразы того же типа, например:

4. *A woggle ugged a diggle,*
5. *An ugg woggles diggs,*
6. *A diggled woggle ugged a woggled diggle.*

Во всех этих случаях вполне разумно говорить о «понимании» текста (оно проявляется в умении определить жанр текста, дать исходные формы слов, поставить вопросы к членам предложения, построить новые правильные фразы). Однако это «понимание» недостаточно для того, чтобы мы могли реагировать на фразы собеседника или предсказать его реакции на такие же фразы, произносимые нами. Если в общем случае предложения языка не обязаны быть осмысленными, диалог (общение, связанный текст) становится невозможным: нельзя эффективно общаться при помощи бессмысленных фраз.

Здесь следует сказать, что описанные выше эксперименты, вообще говоря, недостаточны для вывода о том, что значение не входит в структуру языка. Можно указать на эксперименты, в которых предложения препарировались прямо противоположным образом: в них оставались только корневые морфемы и выкидывались все грамматические, например: *ребен-спа-комнат-шир-распахн-.* Ч. Хоккет, проводивший эти эксперименты <309>¹, отмечает, что им неизменно сопутствовал успех: испытуемые всегда могли предложить толкование предложения, очень часто однозначное. Будучи последовательными, мы должны были бы сделать из них вывод, что в структуру языка входит только значение, но не грамматические правила. Разумеется, этот вывод был бы ложным.

По всей видимости, тексты, не организованные смыслом, рано или поздно должны были бы утратить свою грамматику, а тексты, не организованные грамматическими правилами, должны были бы рано или поздно утратить свой смысл. Если язык действительно является средством общения, то значение, во всяком случае в определенных пределах, входит в структуру языка. Любопытно, что, когда у Р. Якобсона во время его публичной лекции в Московском институте иностранных языков в 1958 году спросили, как он относится к возможности опи-

¹ См. также <158а>.

сать язык, не описывая его значений, он ответил следующей параболой: можно, конечно, отрубить у курицы голову и сделать ценные наблюдения над ее поведением в этом состоянии, но было бы неосторожно утверждать, что для курицы оно является естественным и что, изучив его, мы узнаем о ней все самое существенное.

Попытаемся ответить на вопрос о том, какие типы значений входят в структуру языка и, следовательно, являются предметом лингвистики. В упомянутой на стр. 41 работе И. А. Мельчука о типах языковых значений <12>, <113> различаются, в частности, грамматические и неграмматические значения. Значение называется грамматическим, если в данном языке оно выражается обязательно, т. е. всякий раз, когда в высказывании появляется элемент, значение которого может сочетаться с данным грамматическим значением, причем такие элементы образуют в языке большие классы и поэтому появляются в текстах достаточно часто. Если же некоторое значение выражается не обязательно и не появляется в текстах с достаточно большой частотой, оно считается неграмматическим. С этой точки зрения значение числа в русском языке является грамматическим, так как всякое существительное обязательно имеет показатель числа — единственного или множественного. Грамматические правила русского языка вынуждают нас выражать это значение, независимо от того, считаем ли мы его существенным для сообщения или нет. В противоположность этому в китайском языке значение числа является неграмматическим; если нет нужды специально указать число предметов, о которых идет речь, значение числа остается невыраженным. Аналогичным образом значение вида в славянских языках является грамматическим, а в большинстве романских и германских языков — неграмматическим; в романских и германских языках видовые различия выражаются только в случае необходимости, а в славянских языках — обязательно. Из приведенных примеров следует, между прочим, что значение, являющееся грамматическим в одном языке, может быть неграмматическим в другом.

Не всякое значение может быть грамматическим. Трудно представить себе язык, в котором различие между *иволгой* и *сойкой* выражалось бы грамматически. Обычно в качестве грамматических, т. е. подлежащих обязательному

выражению, выступают более или менее абстрактные значения (времени, числа, деятеля, объекта, причины, цели, контакта, обладания, знания, ощущений, воли, желания, реальности, потенциальности, возможности, умения и т. п.). Как отмечает А. К. Жолковский <110а>, интересными для лингвиста являются те значения, которые хотя бы в некоторых языках являются грамматическими. Следовательно, круг значений, которые должны стать объектом лингвистики, может быть получен, если собрать вместе так называемые грамматические значения самых разных языков; именно эти значения входят в структуру языка (правила кодирования сообщений), независимо от того, являются ли они в данном языке грамматическими или нет.

Изложенный здесь подход к решению важнейшей лингвистической проблемы, кажущийся автору исключительно перспективным, не является, однако, общепринятым. Поэтому в дальнейшем изложении мы учтем и противоположную точку зрения, состоящую в том, что значения не входят в структуру языка.

В зависимости от того, какой смысл вкладывается в понятие «владения языком» (включается ли в него признак владения значением слов или нет), модели речевой деятельности приобретают тот или иной вид и могут быть разделены на: 1) *несемантические*, или чисто синтаксические, имитирующие владение грамматикой, т. е. способность носителей языка понимать и строить грамматически правильные, но не обязательно осмысленные фразы, и 2) *семантические*, которые имитируют способность носителей языка понимать и строить осмысленные предложения.

В зависимости от того, какая сторона речевой деятельности — слушание или говорение — является предметом моделирования, модели речевой деятельности делятся на модели *анализа* и модели *синтеза*. Моделью анализа называется конечное число правил, способных проанализировать бесконечное число предложений данного языка. Синтаксические аналитические модели получают на «входе» текст, а на «выходе» выдают для каждого предложения запись его синтаксической структуры. Семантические аналитические модели получают на «входе» тот же материал, а на «выходе» выдают смысловую запись (изображение смысла) каждого

предложения на специальном семантическом языке. Моделью синтеза называется конечное число правил, способных построить бесконечно большое число правильных предложений. Синтаксические синтетические модели используют в качестве исходной информации запись синтаксической структуры предложений, а на выходе выдают правильные предложения данного языка. Семантические синтетические модели получают на входе смысловую запись некоторого предложения на специальном семантическом языке и выдают на выходе множество предложений естественного языка, синонимичных данному предложению.

Помимо моделей анализа и синтеза, существуют еще так называемые порождающие модели, в некотором смысле промежуточные между моделями анализа и синтеза <199>, <359>. Порождающей моделью называется устройство, содержащее алфавит символов¹ и конечное число правил образования (и преобразования) выражений из элементов этого алфавита, способное построить бесконечное множество правильных предложений данного языка и приписать каждому из них некоторую структурную характеристику.

С понятием аналитических и синтетических моделей связано важное понятие обратимости модели. Модель M_1 называется обратной по отношению к модели M_2 , если исходные объекты M_1 являются конечными объектами M_2 , а конечные объекты M_1 — исходными объектами M_2 . Некоторые исследователи <152> рассматривают синтетические модели как обратные по отношению к аналитическим. При таком подходе к делу отпадает надобность в построении двух самостоятельных моделей данного явления: аналитическая модель может быть получена простым обращением синтетической, и наоборот².

Изобразим схематически отношения между разобранными выше типами моделей (см. схему 2).

В зависимости от того, в какой математической форме излагается модель, модели делятся на исчисления и алгоритмы <5>. Содержательно различие

¹ Под алфавитом символов понимается набор исходных элементов модели, ее словарь — например, набор фонем, или морфем, или словоформ, или символов простейших типов предложений.

² Другое мнение о соотношении анализа и синтеза см. у Р. Якобсона <318, 249>.

между ними можно пояснить следующим образом: исчисление — это система разрешений (позволений), а алгоритм — это последовательность приказов (команд).

Схема 2

Признаки Тип модели	Что известно лингвисту	Характер исходной информации	Характер конечной информации	Цель
Исследовательские	Текст (и множество правильных фраз)	Текст	Грамматика и словарь	Смоделировать деятельность лингвиста
Аналитические	Грамматика и словарь	Текст	Изображение структуры текста	Смоделировать понимание текста
Синтетические	Грамматика и словарь	Изображение структуры текста	Текст	Смоделировать производство текста
Порождающие	Грамматика и словарь	Алфавит символов и правила образования и преобразования фраз	Множество правильных фраз и изображение их структуры	Смоделировать умение отличать правильное от неправильного в языке

Обычно исчисление имеет вид математической системы, включающей: 1) исходные (первичные, или неопределяемые) понятия, имена которых образуют уже знакомый нам «алфавит символов»; 2) первичные (недоказываемые) утверждения о связях между этими понятиями (аксиомы); 3) правила вывода новых утверждений (теорем) из уже имеющихся. Вместо аксиом и правил вывода иногда используются правила образования и преобразования выражений из элементов алфавита. В исчислениях часто используются так называемые рекурсивные определения и правила. Рекурсивными называются опреде-

ления и правила, которые строятся в два шага, причем первый шаг содержит определение простейшего частного случая, а второй — определение общего случая через частный. Примером рекурсивного определения может служить следующее определение натурального числа (т. е. любого целого положительного числа, начиная с единицы): (1) 1 (единица) есть натуральное число; (2) если i — натуральное число, то и $i + 1$ — также натуральное число. Легко убедиться, что под это определение подойдут все натуральные числа, и только они. Пример рекурсивного правила будет приведен ниже¹.

Исчисление позволяет задать с помощью конечного аппарата все объекты некоторого множества, в том числе бесконечного (например, все предложения данного языка). Это свойство исчислений и должно быть использовано лингвистикой, имеющей дело с очень большими или бесконечными инвентарями единиц. В дальнейшем мы изложим несколько лингвистических исчислений; здесь мы проиллюстрируем это понятие одним абстрактным примером.

Дан алфавит символов X, a, c . Обязательное правило образования выражений в этом алфавите следующее: $X \rightarrow ac$ (X переходит в ac). Чтобы наше исчисление порождало бесконечное множество выражений, добавим в него одно факультативное рекурсивное правило: $X \rightarrow aXc$. Следовательно, X переходит либо в ac , либо в aXc , а X в последнем выражении может снова заменяться другими символами по любому из двух правил исчисления. Элемент X , с которым связано это рекурсивное правило, называется рекурсивным. Проиллюстрируем работу построенного здесь исчисления, применив каждое правило по одному разу:

Правило	Результат применения
(1) $X \rightarrow aXc$	aXc
(2) $X \rightarrow ac$	$aacc$

¹ См. рекурсивное определение слова у П. С. Кузнецова <87>, морфемы — у И. И. Ревзина <152>, комплекса и других лингвистических объектов — у С. К. Шаумяна <209>. Ср. также работу И. Бар-Хиллела <18>; в ней ясно и убедительно показано значение рекурсивных определений для эмпирических наук, в том числе лингвистики, которая раньше ими не пользовалась, хотя природа ее материала прямо располагает к этому.

Предлагаем читателю самостоятельно построить хотя бы одно неправильное выражение, не принадлежащее к числу выражений на данном языке (не порожаемое данным исчислением), и указать одно общее свойство всех правильных выражений.

Перейдем теперь к понятию алгоритма. Алгоритмом называется, как мы помним, последовательность команд, выполнение которых ведет к выделению (или построению) желаемого объекта. В качестве иллюстрации рассмотрим алгоритмы решения следующей простой задачи: сложить миллион заданных произвольных чисел, например чисел 12, 1, 102, 29, ..., 5. Простейшим алгоритмом решения этой задачи будет следующий: 1) возьми первое число (12), 2) прибавь к нему второе (1), 3) прибавь к сумме третье (102), 4) прибавь к сумме четвертое (29), ..., 1 000 000 прибавь к сумме миллионное число (5) и выдай результат. Этот алгоритм, содержащий миллион команд, очень непрактичен: бессмысленно повторять миллион раз по существу одно и то же. В хорошем алгоритме стандартная команда должна быть обобщена с помощью одного рекурсивного правила. Мы и попытаемся это сделать. Перенумеруем все числа от первого до миллионного:

12	1	102	29	...	5
1	2	3	4	...	1 000 000

Условимся обозначать буквой i номер произвольного числа (но не само число!). Наконец, уясним себе тот факт, что сумма миллиона чисел формируется постепенно и в начале процесса (до того, как мы взяли первое число) равна нулю. Алгоритм: (1) прими сумму равной нулю; (2) прими i равным единице; (3) прибавь i -е число к сумме; (4) проверь, имеет ли место $i = 1\,000\,000$; (5) да — выдай результат; (6) нет — прибавь к i единицу и делай (3). Этот алгоритм можно представить в виде так называемой блок-схемы (см. рис. 2).

Алгоритм должен допускать совершенно автоматическую реализацию, т. е. реализацию, доступную электронной вычислительной машине; в этом отношении алгоритм можно сравнить с инструкцией для лаборанта, который точно выполняет предписания, проворен, никогда не делает ошибок, но не способен размышлять. Инструкция для такого лаборанта может содержать команды типа приведенных выше, но в ней не должно быть предписаний типа



Рис. 2.

«будь разумен», «поступай правильно», «сделай вывод» или «найди прилагательное» (если нет подробных механически выполняемых правил о том, как это делать) <338>.

Алгоритм, записанный на понятном для машины языке, называется программой.

Любая модель, включая исчисление, должна быть представлена в виде алгоритма (или снабжена алгоритмом), чтобы быть реализованной на машине, потому что машина понимает только язык команд, но не язык разрешений.

Прежде чем перейти к другим типам моделей, укажем на зависимость, существующую между исследовательскими, аналитическими, синтетическими и порождающими моделями, с одной стороны, и алгоритмами и исчислениями — с другой. Первые три типа моделей оформляются обычно в виде алгоритмов, а для изложения порождающих моделей, как правило, используется форма исчислений.

В зависимости от того, какого рода правила используются в модели, различаются вероятностные (статистические) и детерминистские (структурные) модели¹. Существуют и смешанные структурно-статистические модели. Естественные языки в большинстве случаев устроены таким образом, что немногие правила охватывают основное множество фактов, но для объяснения остающихся немногих фактов, большей частью непродуктивных, требуется очень большое число правил. Поэтому в ряде случаев бывает выгодней объяснить данную совокупность фактов не детерминистской моделью, которая из-за обилия правил может оказаться излишне громоздкой для выполнения некоторой вполне определенной задачи, а вероятностной моделью, которая обходится меньшим числом чисто статистических правил и потому менее громоздка. Потеря в точности правил компенсируется в такой модели ее относительной простотой. В качестве иллюстрации можно сослаться на работу И. А. Мельчука об определении рода французского существительного по концу слова <lll>; довольно простые правила позволяют правильно решить этот вопрос в 85 случаях из 100. Для испанского языка аналогичные правила еще более эффективны: они дают правильный ответ в 95 случаях из 100.

Наиболее важными детерминистскими моделями являются модели бинарных дифференциальных структур в области фонологии и морфологии, модель непосредственно составляющих, трансформационная и аппликативная модели в области синтаксиса, модель «семантических множителей» в области семантики.

На этом мы заканчиваем краткий очерк типов исследовательских моделей и моделей речевой деятельности человека. Более подробно они рассматриваются нами в III и IV частях книги соответственно.

Мы ничего не говорим здесь о моделях третьего типа (метамоделях, или метатеории), так как не собираемся

¹ Различные типы чисто статистических моделей (вероятностных и теоретико-информационных) ясно и доступно описаны Р. М. Фрумкиной и Е. В. Падучевой в их работах <12>, <193>. Поэтому мы не будем специально на них останавливаться в последующих частях книги, тем более что они не имеют непосредственного отношения к собственно структурным методам. См также сноску 2 на стр. 122.

описывать их сколько-нибудь подробно. Некоторые элементарные сведения о моделях этого типа, совместимые с жанром данной книги, даются нами в V части.

Глава 3

ЭЛЕМЕНТАРНЫЕ МАТЕМАТИЧЕСКИЕ ПОНЯТИЯ

В этой главе мы познакомим читателя с элементарными математическими понятиями, которые понадобятся нам при изложении материала III, IV и V частей. Заметим, что знание даже этих простейших понятий поможет читателю ориентироваться во многих специальных вопросах.

Рассмотрим прежде всего элементарные представления из области теории множеств. Понятие множества является исходным и не определяется. Множеством может быть названа любая совокупность (класс, собрание) элементов, причем каждый элемент считается ровно один раз. Говорят, что элементы принадлежат множеству (обозначение: $a \in M$). Множество может быть конечным (ср. пример выше) и бесконечным (таково множество чисел $\{1, 2, 3, \dots, n, \dots\}$, образующих натуральный ряд). Оно может содержать один элемент (одноэлементное множество) или не содержать ни одного элемента (пустое множество). Важной характеристикой множества является его мощность; для конечного множества — это просто число его элементов. Число элементов конечного множества, или, иными словами, его мощность, обозначается прямыми скобками: $|M|$ — мощность множества M . Мощность множества $\{a, b, c\}$ равна трем. Пересечением множеств M и N называется такое множество $M \cap N$ (\cap — знак пересечения), элементы которого принадлежат одновременно и M и N . Так, пересечением множеств $\{1, 2, 3\}$ и $\{2, 3, 4, 5\}$ является множество $\{2, 3\}$, а пересечение множеств $\{1, 2, 3\}$ и $\{4, 5, 6, 7\}$ пусто. Графически пересечение обозначается, как показано на рис. 3а. Объединением, или суммой, множеств M и N называется такое множество $M \cup N$ (\cup — знак объединения), каждый элемент которого принадлежит хотя бы одному из этих множеств. Так, объединением множеств $\{1, 2, 3\}$ и $\{2, 3, 4, 5\}$ является множество $\{1, 2, 3, 4, 5\}$, а объединением мно-

жеств $\{1, 2, 3\}$ и $\{4, 5, 6, 7\}$ является множество $\{1, 2, 3, 4, 5, 6, 7\}$. Графически объединение изображается, как показано на рис. 3в. Говорят, что множество M включено в множество N , если каждый элемент, принадлежащий M , одновременно принадлежит и N . Множество $\{1, 3\}$ включено в множество $\{1, 2, 3\}$. Графически включение изображается, как показано на рис. 3с. Каждое включенное множество образует подмножество данного множества. Если подмножества

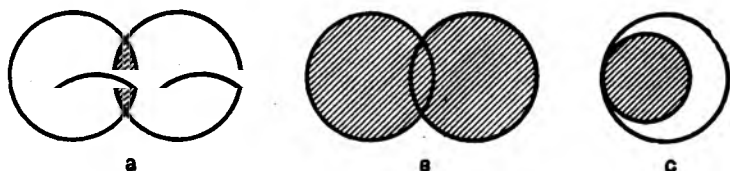


Рис. 3.

данного множества не пересекаются, причем каждый элемент входит в какое-нибудь подмножество, то говорят, что на данном множестве задано разбиение, причем входящие в него подмножества называются классами разбиения. Деление людей на мужчин и женщин является разбиением, а деление людей на знающих иностранные языки и высоких таковым не является, так как часть знающих иностранные языки людей высокого роста, а часть высоких людей знает иностранные языки.

В некотором множестве могут выделяться единички, пары, тройки и вообще n -ки элементов (кортежи длины n). Так, в множестве $\{1, 2, 3\}$ имеются следующие пары (кортежи длины два): $\langle 1, 2 \rangle$, $\langle 2, 3 \rangle$, $\langle 1, 3 \rangle$, $\langle 2, 1 \rangle$, $\langle 3, 2 \rangle$, $\langle 3, 1 \rangle$, $\langle 1, 1 \rangle$, $\langle 2, 2 \rangle$, $\langle 3, 3 \rangle$. Некоторое множество n -ок называется отношением на данном множестве. Особенно важными для лингвиста являются бинарные отношения эквивалентности и порядка; бинарными они называются потому, что в них входят пары элементов. Рассматриваются следующие свойства бинарных отношений (мы будем их иллюстрировать на примере хорошо всем известных отношений «быть равным» и «быть больше»):

1) Рефлексивность. Бинарное отношение R обладает свойством рефлексивности, если имеет место aRa

для любого элемента a , т. е. если любой элемент данного множества находится в отношении R к самому себе. Отношение «быть равным» рефлексивно ($5 = 5$, $2 = 2$ и т. д.), а отношение «быть больше» иррефлексивно (ни для одного a неверно, что $a > a$).

2) **Симметричность.** Бинарное отношение R обладает свойством симметричности, если из aRb следует bRa для любых a и b , входящих в это отношение. Отношение «быть равным» симметрично (из $a = b$ следует $b = a$), а отношение «быть больше» асимметрично (если $a > b$, то заведомо неверно, что $b > a$).

3) **Транзитивность.** Бинарное отношение R обладает свойством транзитивности, если из aRb и bRc следует aRc для любых пар, входящих в это отношение. Отношение «быть равным» транзитивно (из $a = b$ и $b = c$ следует $a = c$). Отношение «быть больше» также транзитивно (из $a > b$ и $b > c$ следует $a > c$).

Бинарное отношение R называется эквивалентностью, если оно обладает свойствами рефлексивности, симметричности и транзитивности. Равенство есть частный случай эквивалентности. Бинарное отношение R называется отношением строгого порядка, если оно иррефлексивно, асимметрично и транзитивно. Таково отношение «быть больше». Важными являются и отношения частичного порядка, которые рефлексивны, симметричны только в случае равенства a и b и транзитивны. Примером частичной упорядоченности является отношение «больше или равно» ($a \geq b$).

Существует интересная связь между разбиениями и эквивалентностями, состоящая в том, что каждому разбиению данного множества соответствует некоторое отношение эквивалентности, а каждое отношение эквивалентности на данном множестве порождает некоторое его разбиение.

Теперь покажем в самых общих чертах (подробности читатель найдет в III, IV и V частях книги), каким образом введенные здесь понятия могут быть использованы лингвистом. В лингвистике мы имеем дело с множеством элементов (звуков, фонем, морфем, словоформ, слов, словосочетаний, предложений), причем: 1) языковые единицы довольно часто определяются как классы эквивалентных единиц текста (фонемы — как классы звуков, морфемы — как классы морфов, слова —

как классы словоформ, значения — как классы употреблений); 2) языковые единицы к л а с с и ф и ц и р у ю т с я (фонемы — на гласные и согласные, слова — на части речи и т. д.), а всякая корректно построенная классификация должна представлять собой р а з б и е н и е; 3) языковые единицы внутри класса и классы внутри классификации часто представляются в виде «иерархии» (ср. «главный» член парадигмы, «исходное» слово словообразовательного гнезда, «основные» части речи, «главные» и «второстепенные» члены предложения), а всякая иерархия предполагает некоторое отношение п о р я д к а.

Отчетливое понимание своих действий при решении названных здесь и других аналогичных вопросов поможет лингвисту избежать неточностей и ошибок, которые в противном случае будут неизбежны.

Рассмотрим теперь некоторые элементарные понятия теории графов, находящей все более широкое применение в самых различных областях лингвистики, в особенности в синтаксисе, где основные объекты этой теории — графы и деревья — используются для описания синтаксической структуры предложений.

Г р а ф о м называется пара множеств, одно из которых (ребра, ср. отрезки ab , ac , ad , ae , af , bf , be , bd , bc , cf , ce , cd , df , de , ef на рис. 4 (2)) находится в одно-двузначном соответствии с другим (вершины, ср. точки a , b , c , d , e , f на рис. 4 (2)); каждому ребру соответствуют две вершины. Примеры графов см. на рис. 4, стр. 117.

Графы, в которых любая пара точек соединена ребром, называются п о л н ы м и. На рисунке 4 полными являются графы (1) и (2), а все остальные — н е п о л н ы м и. Ломаная линия¹, соединяющая какие-нибудь две вершины графа, называется п у т е м из одной вершины в другую. Путь, начало которого совпадает с концом, образует ц и к л. На нашем рисунке графами с циклами являются (1), (2) и (3). В частности, в графе (2) цикл образует любая тройка, четверка, пятерка и шестерка точек. Граф, в котором из любой точки можно пройти в любую другую точку, называется с в я з н ы м; если это условие не выполнено, то граф называется н е с в я з н ы м. На нашем рисунке все графы, за исключением (5), связные. Связный граф, не содержащий циклов, называется д е р е в о м. На нашем

¹ Следует иметь в виду, что прямая — частный случай ломаной.

рисунке деревом является только граф (4); остальные графы деревьями не являются. Ребра дерева иногда называются ветвями, а вершины — узлами. Ребра графа можно ориентировать, показав, в каком направлении следует переходить из одной точки в другую; для этого достаточно нарисовать стрелку на одном из концов ребра. Граф с ориентированными ребрами называется **о р и е н т и р о в а н н ы м**; на нашем рисунке ориентирован граф (1), а все остальные **н е о р и е н т и р о в а н ы**.

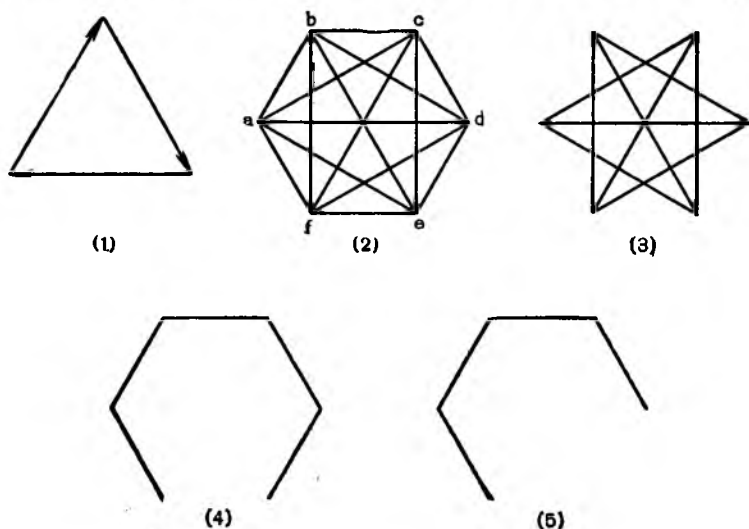


Рис. 4

Последняя область математики, понятия которой будут использоваться нами в дальнейшем, — это теория вероятностей.

Пусть дано некоторое **п о л е с о б ы т и й**, например выпадение чисел 1, 2, 3, 4, 5 или 6 при бросании игральной кости. Бросание кости назовем **и с ы т а н и е м**, а выпадение того или иного из 6 чисел — **и с х о д о м** испытания, или **с о б ы т и е м**. **В е р о я т н о с т ь ю** события называется отношение числа благоприятных для данного события исходов m к числу всех возможных исходов n . Вероятность каждого из наших событий, т. е. вероятность

того, что при бросании игральной кости (которую мы предполагаем совершенно симметричной) выпадет какое-то из шести чисел, равна $\frac{1}{6}$. Если события независимы друг от друга, то их вероятности в сумме дают единицу. Минимальное возможное значение вероятности равно нулю, а максимальное возможное значение — единице. Условной вероятностью события A при условии B называется вероятность того, что произойдет A , при условии, что B уже имеет место. Условная вероятность вычисляется по формуле

$$P(A/B) = \frac{P(A \cdot B)}{P(B)},$$

где $P(A/B)$ обозначает вероятность A при условии B , $P(A \cdot B)$ обозначает вероятность совместного появления двух событий A и B , а $P(B)$ — вероятность события B (вероятность условия). Чтобы освоиться с понятием условной вероятности, решим следующую задачу: найти вероятность выпадения двойки (числа 2), при условии что уже выпало четное число. Найдем сначала вероятность совместного появления двух событий — выпадения двойки и четного числа одновременно. Она равна $\frac{1}{6}$. Вероятность выпадения четного числа (вероятность условия) равна $\frac{1}{2}$ (игральная кость имеет три четных числа и три нечетных; вероятность того, что выпадет четное число, равна $\frac{3}{6} = \frac{1}{2}$). Следовательно, вероятность выпадения двойки, при условии что уже выпало четное число, равна $\frac{1}{6} : \frac{1}{2} = \frac{1}{3}$.

В экспериментальных условиях мы имеем дело не с вероятностями, а с относительными частотами событий. Однако относительные частоты приближаются к вероятностям по мере увеличения числа испытаний.

Изложенные выше элементарные понятия теории вероятностей используются во многих исследовательских и некоторых порождающих моделях, рассматриваемых нами ниже.

Некоторые другие математические понятия, имеющие более частный характер, излагаются по ходу дела в соответствующих частях книги.

МОДЕЛИ ИССЛЕДОВАНИЯ

В главе 2 предыдущей части мы сказали, что модели, имитирующие исследовательскую деятельность лингвиста, предназначены для того, чтобы объективно обосновать используемые в лингвистике понятия. Это необходимо по двум причинам. Во-первых, далеко не для всех лингвистических утверждений можно на нынешнем этапе развития науки указать решающие эксперименты, которые могли бы их подтвердить или опровергнуть. В подобном положении находится, например, удивительно красивая «гипотеза лингвистической относительности» Э. Сэпира и Б. Уорфа, изложенная нами в I части книги. В этих условиях объективное обоснование научных понятий и утверждений становится едва ли не единственным способом убедиться в их правильности.

Во-вторых, существуют и более глубокие соображения в пользу того мнения, что исследовательские модели, в которых научным понятиям дается строгое обоснование, представляют принципиальный интерес. В науке существует так называемый принцип предшествования: биологии предшествует химия, и понятия, которые биология принимает на веру, тщательно обосновываются химиком; химии предшествует физика, в частности физика атома, и физик ищет обоснование для тех понятий, которые химик использует в качестве исходных и неопределяемых. В свое время И. П. Павлов много сделал для создания экспериментальной психологии, которой предшествует физиология: простейшие проявления психической деятельности человека он пытался выразить в физиологических понятиях, оставаясь в роли «экспериментатора, имеющего дело исключительно с внешними явлениями и их отношениями» (188, 55). Наиболее точным эмпири-

ческим наукам предшествует математика, оперирующая весьма простыми и общими понятиями. Тем не менее даже математик не принимает на веру свои собственные понятия, не исключая такие, казалось бы, очевидные, как понятие числа. В математике существует особая дисциплина — основания математики, — в задачи которой входит, в частности, конструктивное выведение простых понятий из простейших.

Из сказанного следует, что одна из важных задач каждой эмпирической науки заключается в сведении используемых в ней понятий к понятиям той науки, которая ей предшествует. Только на этом пути возможно создание единой научной картины мира, в которой нет оборванных концов.

Как и другим стремящимся к точности эмпирическим наукам, структурной лингвистике в конечном счете предшествует математика. Объективно обосновать лингвистические понятия — значит найти опирающуюся на математические понятия конструктивную процедуру, ведущую к их выделению на основе минимальной заранее заданной информации о языке. Исследовательские модели и представляют собой, в идеале, наборы таких процедур.

Глава 1

МОДЕЛИ ДЕШИФРОВКИ

Первый тип исследовательских моделей, который будет нами рассмотрен, можно условно назвать *моделями лингвистической дешифровки*, так как, помимо своей основной теоретической функции (обоснования лингвистических понятий и утверждений), они в принципе могут иметь и некоторую прикладную — дешифровочную — функцию. Как мы помним, исходной информацией для моделей этого типа является текст, о котором заранее ничего не известно. Неизвестны ни язык (код), использованный для «шифровки» текста, ни генетические связи этого языка с уже известными языками, ни переводы текста на известные языки, ни та область действительности, которая описывается текстом. Для наглядности можно представить, что к нам в руки попал текст на марсианском языке, описывающий незнакомую нам

марсианскую действительность и не имеющий никаких связей с текстами на известных нам языках. Помимо самого текста, мы имеем право использовать для его дешифровки только наше умение отличать черные точки от белых (этому легко «научить» и электронную вычислительную машину). Все остальные сведения, т. е. сведения об элементарных единицах текста (буквах или звуках, морфемах, словах, предложениях и, наконец, смыслах), классах элементарных единиц (гласных и согласных, лексических и грамматических морфемах, частях речи, типах предложений, семантических полях) и законах сочетания единиц различных классов (например, синтаксических связях слов в предложении), должны быть совершенно автоматически получены из текста.

Можно представить себе последовательность алгоритмов, в которой каждый алгоритм выполняет одну из названных выше задач. Входной информацией для первого алгоритма являются текст и сведения о черных и белых точках (умение отличить черное от белого), а на выходе он вырабатывает информацию об алфавите символов (например, букв), с помощью которого этот текст записан. Каждый последующий алгоритм, решая ту или иную задачу, получает на вход информацию, выработанную предыдущим алгоритмом. В частности, второй алгоритм находит в алфавите символов, который был обнаружен первым алгоритмом, гласные и согласные, подклассы внутри гласных и согласных и т. д., пока не будет установлено чтение всех букв. Последующие алгоритмы, пользуясь этой информацией, находят слоги, а затем морфемы, слова, классы морфем и классы слов. Когда найдены классы слов, можно приступить к решению синтаксических задач, в частности установить границы предложений и обнаружить связи слов в предложении. Наконец, когда открыты все существенные черты грамматики, можно переходить к поиску смысла слов и предложений. Результатом работы алгоритмов должно быть такое представление о языке, которое достаточно для того, чтобы перевести изученные таким образом тексты на какой-либо уже известный язык или сопоставить текстам изображенный в них кусок действительности.

Если эта программа, впервые научно поставленная в пионерских работах Б. В. Сухотина <173>, <174>, окажется реализуемой, то основные лингвистические понятия будут сведены к весьма простым понятиям черных и

белых точек, ряду содержательных гипотез и описывающим их математическим функциям. Даже если она нереализуема, интересно выяснить, до какого уровня такого рода описание может быть успешно доведено.

Ниже мы излагаем наиболее простые алгоритмы Б. В. Сухотина, З. Харриса и ряда других исследователей, которые можно отнести к числу дешифровочных¹. Укажем некоторые общие черты всех этих алгоритмов. Во-первых, в основе всех алгоритмов лежат простые и общие представления о языке, подтверждающиеся определенными универсальными закономерностями, например: «буква есть устойчивое сочетание точек»; «морфема есть устойчивое сочетание фонем»; «словоформа есть устойчивое сочетание морфем»; «в каждом естественном языке имеется минимум два уровня — уровень значащих единиц (морфем, словоформ, конструкций) и уровень незначащих единиц (фонем)»; «в любом языке имеются лексические морфемы, причем распределение лексических морфем в тексте отличается от распределения грамматических морфем»; «синтаксическим различиям соответствуют семантические различия»; «слова, близкие по смыслу, стоят в тексте недалеко друг от друга» и т. п. Во-вторых, во всех алгоритмах такого рода используется информация о дистрибуции элементов и их числовых параметрах². В-третьих, задача обычно решается следующим образом. Сначала определяется множество допустимых решений, а затем в этом множестве с помощью так называемых функций выгодности находится наилучшее решение. Функциями выгодности называются числовые функции, которые в случае правильных (наилучших) решений принимают определенное (например, минимальное возможное или максимальное возможное) значение. Каждая функция выгодности формализует некоторую содержательную гипотезу о возможных свойствах искомого объекта. Впервые, правда, в очень нестрогой форме, функции выгодности были использованы

¹ К обсуждаемому здесь вопросу имеют отношение работы Н. Д. Андреева <3>, <4>, Х. Спанг-Ханссена <358>, Ю. А. Шрейдера <221>.

² О соотношении структурных (в том числе дистрибутивных) и числовых (в том числе статистических) методов см. <111>, <177>, <148>, <193>, <199>, <323>, <287>, <295>, <293>, <195>, <305>, <357>, <358>, <363>, <376> и др.

Л. Ельмслевом <129, 171> и З. Харрисом <295, 63>, и мы воспользуемся их высказываниями, чтобы пояснить это понятие. Рассмотрим следующий пример. Дано множество звуков. Требуется разбить его на классы, соответствующие фонемам. Допустимым решением можно считать произвольное разбиение множества звуков на классы. Содержательная гипотеза Л. Ельмслева и З. Харриса о тех свойствах фонем, благодаря которым мы можем их обнаружить в множестве звуков, состоит в том, что фонемы относительно немногочисленны, но зато каждая из них встречается в большом числе окружений. Следовательно, число фонем (классов звуков) в данной классификации и число окружений для каждой фонемы могут служить теми функциями выгоды, с помощью которых в множестве допустимых решений отыскивается правильное. Правильному решению, как следует из замечаний Л. Ельмслева и З. Харриса, соответствует небольшое значение первой функции и большое значение второй.

Начнем изложение упомянутых выше алгоритмов с алгоритма выделения гласных и согласных¹ Б. В. Сухотина. Этот алгоритм предполагает, что уже выделен алфавит букв, с помощью которого записан данный текст. Его задача состоит в том, чтобы дать нам ключ к правилам чтения текста, без которых дешифровка не может считаться законченной. Допустимым решением считается любое разбиение алфавита на два класса. Алгоритм формализует следующую содержательную гипотезу о свойствах гласных и согласных букв: гласные и согласные в тексте чередуются; не существует текстов, состоящих из одних гласных, как не существует текстов, состоящих из одних согласных; за гласными следуют согласные, а за согласными—гласные, ср. *корова, палата, собака* и т.п. Однако эта закономерность не является строгой; возможны скопления гласных и согласных, ср. *театр, клоака, страх, встроенный* и т. п. Следовательно, максимум того, на что мы можем надеяться,— это преобладание чередований гласных и согласных по сравнению со скоплениями: рядом с гласными чаще стоят соглас-

¹ Речь идет не о гласных и согласных звуках, а о гласных и согласных буквах. Однако все, что говорится ниже о текстах, записанных буквами, сохраняет силу и для текстов, записанных фонетической транскрипцией.

ные, чем гласные, а рядом с согласными чаще стоят гласные, чем согласные. Самая частая буква текста — гласная (закономерность, подтверждаемая материалом многих языков).

Составим квадратную таблицу частот двухбуквенных сочетаний; строки и столбцы таблицы озаглавим буквами алфавита (табл. 1); в каждой клетке проставим число сочетаний буквы a_i с буквой a_j , встретившихся в данном тексте (т. е. число сочетаний a_1 с a_2 , a_3 и т. д.); на нашем рисунке каждое сочетание изображается точкой.

Таблица 1

	a_1	a_2	a_3	a_4	a_5	
a_1	11 5 — 3
a_2	12 4 — 2
a_3	13 9
a_4	14
a_5	10 4 — 4

Таблица 2
Гласные Согласные

Г л а с н ы е	1	2
	4	3

Допустим, что мы уже расклассифицировали гласные и согласные, причем гласные отодвинуты в левый верхний угол таблицы (табл. 2). Если наша гипотеза верна, то указанная выше общая закономерность строения текстов отразится на такой таблице следующим образом: число отметок в квадратах 1 и 3 будет небольшим (потому что гласные редко сочетаются с гласными, а согласные редко сочетаются с согласными), а число отметок в квадратах 2 и 4 будет гораздо большим (потому что гласные часто сочетаются с согласными, а согласные — с гласными). Следовательно, чтобы найти правильное решение нашей задачи в множестве допустимых решений, необходимо преобразовать исходную таблицу в таблицу указанного

здесь вида (табл. 2), т. е. переставить строчки и столбцы таким образом, чтобы сумма отметок в диагональных квадратах 1 и 3 была наименьшей.

Этой цели служит следующий алгоритм:

(1) Найди сумму отметок для каждой строки.

(2) Проверь, имеются ли строки с положительными суммами. Нет — выдай результат. Да — делай (3).

(3) Найди строку с максимальной суммой отметок и занеси букву, которая озаглавливает эту строку, в класс гласных.

(4) Найди столбец, озаглавленный выделенной в (3) буквой, и вычти из суммы отметок для каждой строки, за исключением уже выделенных, удвоенное число, стоящее на пересечении данной строки и столбца, озаглавленного выделенной в (3) буквой.

(5) Сотри предыдущие суммы отметок в строках и делай (2).

Обработаем этим алгоритмом таблицу 1. Выполнение первой команды дает результаты (11, 12, 13, 14 и 10), записанные в первой колонке направо от таблицы. Выполняем вторую команду и убеждаемся, что все суммы положительны. Строкой с максимальной суммой, которую мы должны найти в результате (3), является четвертая строка (сумма отметок в ней равна 14). Заносим a_4 в класс гласных. Результаты выполнения команды (4) представлены во второй колонке справа от таблицы; они найдены следующим образом: $11 - 2 \cdot 3 = 5$ для первой строки, $12 - 2 \cdot 4 = 4$ для второй, $13 - 2 \cdot 2 = 9$ для третьей и $10 - 2 \cdot 3 = 4$ для пятой. Стираем, по (5), старые суммы и возвращаемся к (2). Убеждаемся, что все суммы (5, 4, 9, 4) положительны, выбираем строку с максимальной суммой (третью) и заносим букву a_3 в класс гласных. Выполняем все остальные команды и убеждаемся, что положительных сумм не остается (ср. числа -3 , -2 и -4 в третьей колонке). Следовательно, a_3 и a_4 входят в класс гласных, а a_1 , a_2 и a_5 — в класс согласных. Легко убедиться, что в результате мы получили таблицу требуемого вида¹ (табл. 3, стр. 126).

Изложенный алгоритм Б. В. Сухотина был опробован в ряде машинных экспериментов, проведенных на

¹ В общем случае, как показал Б. В. Сухотин, этот алгоритм не находит абсолютного минимума; однако здесь было бы нецелесообразно обсуждать это осложнение и способы его снятия.

материале русского, английского, французского, немецкого и испанского языков (на каждом языке был

Таблица 3

взят текст длиной в 10 000 знаков). На немецком материале алгоритм дал 3 ошибки (*s*, *h* и *k* попали в число гласных, очевидно, из-за частотности сочетаний *sch*, *ch*, *ck*); на английском, французском и русском материале алгоритм дал всего по одной незначительной и легко объяснимой ошибке; на испанском материале он проработал безошибочно¹.

	a_3	a_4	a_1	a_2	a_5
a_3
a_4
a_1
a_2
a_5

Изложенный здесь алгоритм, разумеется, не является достаточной основой для перехода на морфологический уровень. Необходимо иметь еще алгоритм выделения слов, отличный от алгоритма выделения морфем, алгоритм перевода слоговой письменности в буквенную и много других. Некоторые из этих алгоритмов уже разработаны Б. В. Сухотиным, но мы их не будем здесь излагать ввиду их сложности; по той же причине мы вынуждены будем отказаться от рассмотрения весьма интересного морфологического алгоритма Б. В. Сухотина <174> и изложим вместо него более простой алгоритм выделения морфем (точнее морфов), принадлежащий З. Харрису <294>.

Алгоритм З. Харриса находит не морфы, как таковые, а границы между ними. Чтобы понять основную идею этого алгоритма, рассмотрим популярную игру в «балду», которая хорошо ее иллюстрирует. В этой игре какой-нибудь из участников пишет первую букву задуманного им слова (существительного). Каждый следующий участник добавляет справа к уже написанным буквам новую букву, стараясь

¹ Аналогичные идеи и приемы были использованы В. В. Шеворошкиным в блестяще удавшейся ему дешифровке карийского языка (см. В. В. Шеворошкин, Карийский язык, изд. «Наука», М., 1965).

повернуть игру таким образом, чтобы конец слова не пришелся на него. Проигравшим считается тот участник, который вынужден закончить слово, а чей-то проигрыш неизбежен потому, что бесконечных слов нет, и по мере продвижения от начала слова к его концу возможности повернуть игру в новое русло, т. е. изменить строящееся слово, непрерывно сокращаются: та часть слова, которая уже построена, все с большей и большей определенностью предсказывает его конец. Однако в конце слова (за его последней буквой) снова происходит «всплеск» неопределенности, потому что начинать новое слово можно почти любой буквой алфавита. Следовательно, границам слов соответствуют «всплески» (пики) неопределенности ¹.

Похожая мысль была положена в основу морфологического алгоритма З. Харриса. Алгоритм работает следующим образом. В тексте, записанном фонологической транскрипцией, выбирается некоторое предложение, например англ. [hiyzklevar]² — «он умен». Отыскиваются все предложения, начинающиеся с той же фонемы [h], и подсчитывается число различных фонем, которые следуют за [h] в этих предложениях. Эти фонемы называются «преемниками» (successors) [h]. Число «преемников» для [h] равно 9. Затем отыскиваются все предложения, которые начинаются с тех же двух фонем, что и данное, т. е. с [hi], и подсчитывается число «преемников» для пары фонем (оно равно 14) и т. д. В разных местах предложения число «преемников» изменяется: то возрастает, образуя пики, то падает. Если верно предположение, что фонологическое разнообразие на стыке морфов больше, чем внутри морфов, то пики должны приходиться на границы между морфами, ср.

h	i	y	z	k	l	e	v	ə	r
9	14	29	29	11	7

¹ Сведущий читатель заметит, что эти рассуждения можно было бы провести в теоретико-информационных терминах; заслуживает внимания то обстоятельство, что в машинных экспериментах (см. ниже) действительно подсчитывалась энтропия.

² Транскрипция принадлежит З. Харрису.

В данном предложении имеются два пика — после фонемы [y] (29) и после фонемы [z] (29); в этих местах и следует провести морфологические границы. Полученная сегментация (*he/'s/clever*) отражает действительное положение вещей. Интересно, что для фразы *he's quicker* — «он более проворен», внешне очень похожей на фразу *he's clever*, получается иная, и тоже правильная, сегментация — с еще одним пиком после фонемы [k]: третья граница проходит перед морфемой сравнительной степени *-er*.

Для устранения возможных ошибок вводится еще несколько правил, из которых мы приведем здесь только одно: подсчитывается не только число «преемников» фонем при движении от начала предложения к его концу, но и число «предшественников» (*predecessors*) фонем при движении от конца предложения к его началу.

Упомянем в заключение результаты машинных экспериментов <12, 141> по проверке предложенного З. Харрисом метода морфологической сегментации текста. Эти эксперименты, проведенные на материале 100 английских предложений в фонологической транскрипции, дали вполне удовлетворительные результаты: в 85% случаев машина правильно сегментировала текст.

Мы уже сказали, что алгоритм такого типа выделяет, вообще говоря, не морфемы, а морфы, которые должны быть объединены в морфемы на основании каких-то других принципов. Предложенный З. Харрисом метод морфологического анализа наталкивается и на другие трудности, многие из которых были проанализированы им самим (метод не дает возможности выделить так называемые разрывные и другие несегментные морфемы, суперсегментные морфемы и т. п.). Со своей стороны, мы бы хотели обратить внимание на еще одну специфическую трудность, связанную с наличием в языке всякого рода слитных и наложенных морфем.

Одно из важнейших допущений рассматриваемого метода и других подобных ему методов состоит в том, что морфемы в тексте не пересекаются, т. е. не бывает случаев, когда конец одной морфемы является началом другой (как в чайнвордах). В большинстве случаев тексты действительно построены таким образом, но общий случай более сложен. Весьма показательны в этом отноше-

нии соответствующие факты русского языка, описанные под общим названием «наложения морфем» (частный случай гаплогонии) Е. А. Земской <67>. В проанализированных Е. А. Земской примерах наложение морфем (точнее, морфов) имеет место в тех случаях, когда последние фонемы основы совпадают с первыми фонемами словообразовательного суффикса, ср. *дербист*, *таксист* в противоположность *футбол-ист*: в словах *дербист*, *таксист* гласный *и* принадлежит и основе, и суффиксу (ср. более «правильные», хотя и несуществующие формы *дерби-ист*, *такси-ист*). Другими примерами такого рода являются слова *лиловатый* (в противоположность *черн-оватый*), *Мэрин* (в противоположность *Кат-ин*), *сальеризм* (в противоположность *вагнер-изм*).

В этих и других подобных случаях алгоритм З. Харриса будет делать ошибки, но из этого отнюдь не следует, что не существует методов коррекции, способных их устранить.

Получение морфологического членения текста является лишь первым, хотя и наиболее важным шагом на пути к получению его грамматики. После этого необходимо, как минимум, 1) отделить лексические морфемы (корни) от грамматических; 2) обнаружить классы корней (например, основы производных существительных, производных глаголов, производных прилагательных и т. п.) и классы грамматических морфем (парадигмы), 3) научиться находить словоформы и классы эквивалентных словоформ, 4) научиться находить границы предложений и правила синтаксической связи между словоформами в составе предложения. Само собой разумеется, что каждая из этих задач должна решаться особым алгоритмом.

Ниже мы изложим лишь алгоритм установления синтаксических связей словоформ в предложении, принадлежащий Б. В. Сухотину <174>; однако изложению этого алгоритма мы предположим несколько общих соображений о том, каким образом может быть решена задача отделения лексических морфем от грамматических. Этой задачей длительное время занимался А. Жюйян <323>, рассматривавший два числовых параметра морфем: 1) их частоту в текстах, 2) их распределение в текстах различных жанров. Он пришел к выводу, что высокая частота встречаемости при ровном распределении в самых различных текстах характерна для грамматических морфем, а не-

высокая частота в соединении с неравномерным распределением в текстах характерна для лексических морфем, в особенности терминов. Эти соображения кажутся довольно очевидными: морфемы именительного и винительного падежей, например, встречаются с более или менее неизменной и весьма высокой частотой в текстах самых различных жанров, в то время как корневые морфемы слов типа *интеграл*, *логарифм*, *синус* и т. п. не только гораздо менее частотны, но и неравномерно распределены (встречаются главным образом в математических текстах). Поэтому изученные А. Жюйяном параметры могут стать основой функций выгодности, предназначенных для решения названной здесь задачи.

Другой путь решения этой задачи был указан Б. В. Сухотиным. Он обратил внимание на то, что грамматические и лексические морфемы ведут себя в некоторых отношениях подобно гласным и согласным: в тексте они чередуются (за группой лексических морфем следует группа грамматических, и наоборот), причем наиболее частая морфема всегда грамматическая. Поэтому для обнаружения в множестве морфем класса лексических морфем и класса грамматических морфем можно использовать изложенный выше алгоритм выделения гласных и согласных (или вариант этого алгоритма, с поправкой на специфику морфологического материала).

Сходные идеи, высказанные Ю. В. Кнорозовым, легли в основу практической работы по дешифровке непрочтенных письменностей, проводимой во Всесоюзном институте научной и технической информации АН СССР. По мысли Ю. В. Кнорозова, в класс лексических морфем входит б о л ь ш о е число единиц, каждая из которых сочетается с н е б о л ь ш и м числом единиц другого класса, т. е. класса грамматических морфем; в этот последний входит н е б о л ь ш о е число единиц, но зато каждая из них сочетается с б о л ь ш и м числом единиц другого класса, т. е. класса лексических морфем.

Перейдем к алгоритму установления синтаксических связей словоформ в предложении. Он распадается на две части, из которых первая представляет собой модель исследования в собственном смысле слова: ее выходом являются элементарные синтаксические правила данного языка. Во второй части алгоритма, выходящей за рамки собственно исследовательской модели, эти пра-

вила применяются для анализа синтаксической структуры предложения и получения его дерева.

Алгоритм работает над текстом, словоформы которого заменены (перекодированы) символами классов словоформ, предполагаемых заранее заданными. Используются следующие символы: 1) N_n — существительные и личные местоимения в именительном падеже; 2) N_g, N_d, N_a, N_i, N_p — то же самое соответственно в родительном, дательном, винительном, творительном и предложном падежах; 3) A_n, A_g, A_d, \dots — прилагательное и его эквиваленты в именительном, родительном, дательном и т. д. падежах; 4) V — глагол в личной форме; 5) P — предлог; 6) C — союз и т. д.

По указанным здесь правилам перекодирования словоформы предложения *В окна брезжил синеватый холодный свет утра; под лавкой шипел и крякал проснувшийся селезень* будут перекодированы следующим образом: $PN_aVA_nA_nN_nN_g; PN_iVCVA_nN_n$. Алгоритм устанавливает связи словоформ в простом предложении; поэтому до начала работы этого алгоритма какой-то другой алгоритм должен разбить каждое сложное предложение на простые, не содержащие обращений, вводных слов и т. п. Для каждого простого предложения должна быть получена схема связей типа тех, которые каждый из нас рисовал в школе;

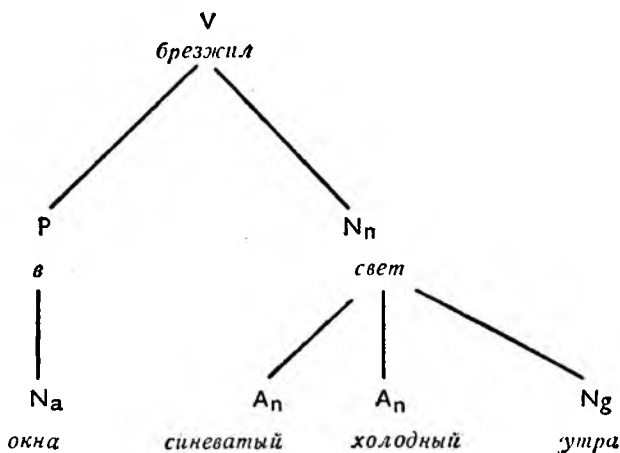



Рис. 5.

в частности, для предложения *В окна брезжил синеватый холодный свет утра* схема должна выглядеть, как на рис. 5.

Алгоритм формализует следующую содержательную гипотезу о существовании синтаксических связей: пары словоформ, непосредственно синтаксически связанных друг с другом, принадлежат к таким классам словоформ, элементы которых часто встречаются в составе одного и того же предложения. Считается, что порядок следования словоформ не является существенным: если словоформы стоят рядом, это еще не значит, что они синтаксически связаны; с другой стороны, если они находятся на значительном удалении друг от друга, это еще не значит, что они синтаксически не связаны.

Для изложения алгоритма нам понадобятся некоторые понятия теории графов, рассмотренные нами на стр. 116.

Легко заметить, что приведенная выше схема предложения отвечает определению дерева. Следовательно, в результате работы алгоритма мы должны получить дерево предложения — связный граф, в котором нет циклов. Граф должен быть связным, так как в простом предложении любая словоформа непосредственно синтаксически связана хотя бы с одной другой словоформой¹. Требование о том, что граф не должен содержать циклов, кажется менее очевидным: в предложениях с субъектным или объектным предикативом форма именной части зависит и от глагола (по управлению), и от подлежащего или дополнения (по согласованию); в частности, падеж именной части определяется глаголом, а род и число — существительным, ср. *Они стали взрослыми; Он стал взрослым; Она стала взрослой* или *Он считает ее безнравственной; Она считает его безнравственным* и т. п. Если изображать в графе предложения не только связи управления, но и связи согласования, то в нем получается цикл, ср.

*Они стали взрослыми* Однако по некоторым серьезным причинам, обсуждение которых увело бы нас в сторону от основной темы, связи согласования в таких случаях не указываются.

¹ Этому предположению противоречит как будто существование обстоятельств, относящихся ко всему предложению; однако обстоятельства, относящиеся ко всему предложению, можно считать связанными с его вершиной, т. е. со сказуемым.

Исходной точкой работы алгоритма является полный граф предложения (рис. 6). Задача состоит в том, чтобы выкинуть из полного графа все «лишние» ребра, оставив только те ребра, которые соответствуют непосредственным синтаксическим связям словоформ в данном предложении. Иными словами, в множестве допустимых решений (полном графе) необходимо выбрать наилучшее решение. Понять задачу в этой постановке поможет образ фильтра, который получает на «входе» массу ненужного материала, но пропускает на «выход» только то, за чем мы охотимся.

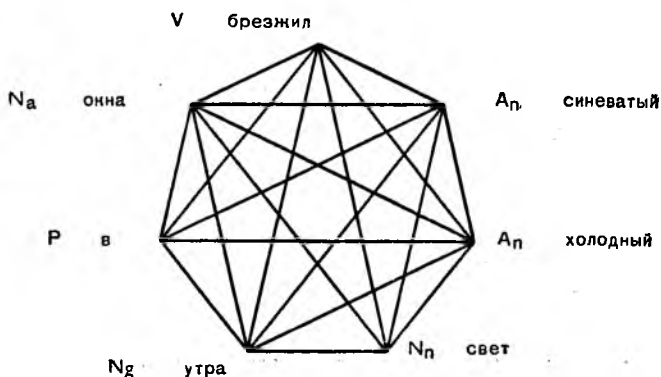


Рис. 6.

Как сказано выше, алгоритм работает в два этапа. Сначала для данного текста составляется «грамматика», имеющая вид квадратной таблицы. Столбцы и строки таблицы озаглавлены символами классов словоформ (см. табл. 4, стр. 134).

Таблица заполняется следующим образом (излагается упрощенный вариант): берется первое предложение текста и отыскиваются все встречающиеся в нем пары словоформ. Для нашего предложения — это пары {в, окна}, {в, брезжил}, {в, синеватый}, {в, холодный}, {в, свет}, {в, утра}, {окна, в}, {окна, брезжил}, {окна, синеватый}, {окна, холодный}, {окна, свет}, {окна, утра} и т. д., всего 30 пар. В таблице отыскивается строка, заглавие которой соответствует первой словоформе данной пары, и столбец, заглавие которого соответствует второй словоформе пары. Для первой пары — это строка P и столбец N_a, для второй — та же строчка

ка и столбец V , для третьей — та же строчка и столбец A_n и т. д. В клетку, стоящую на пересечении строчки и столбца, вносится единица, независимо от того, встретились ли пара словоформ данного вида один раз или несколько.

Таблица 4

	N_n	N_g	N_a	A_n	V	P	...
N_n		1	1	1	1	1	...
N_g	1		1	1	1	1	...
N_a	1	1		1	1	1	...
A_n	1	1	1		1	1	...
V	1	1	1	1		1	...
P	1	1	1	1	1		...
...

После того как по этим правилам обработано первое предложение (на нашей табличке представлен именно этот результат), аналогичным образом обрабатывается второе и т. д., пока не будет охвачен весь текст, для которого составляется грамматика. К концу работы клетки таблицы оказываются заполненными числами, причем в одних клетках числа больше, в других — меньше. Очевидно, например, что число в клетке N_nV будет больше числа в клетке A_nV , так как, если в предложении встречается глагол в личной форме, весьма велика вероятность того, что оно будет содержать и существительное в именительном падеже, в то время как прилагательное в именительном падеже в таком предложении вполне может отсутствовать, ср. *Часы бьют, Брезжит день, Воют сирены* и т. п. Число, стоящее в клетке N_nA_n , будет больше числа, стоящего в клетке N_nN_a , так как,

если в предложении встречается прилагательное в именительном падеже, в нем почти обязательно должно быть и существительное в именительном падеже, в то время как наличие существительного в винительном падеже не требует существительного в именительном падеже, ср. *Длинный состав медленно полз по равнине; Яркое солнце лениво подымалось над лесом и Прочитай эту книгу; Пшеницу побило градом; Его знобило.*

Итак, на первом этапе решается задача нахождения грамматики языка по текстовым данным: составленная нами таблица содержит, в вероятностной форме, все правила связи словоформ в предложении. Имея эту вероятностную грамматику, мы возвращаемся, на втором этапе работы алгоритма, к первому предложению текста, составляем полный граф этого предложения (рис. 6) и над каждым ребром графа пишем цифру, взятую из соответствующей клетки таблицы (поскольку таблица симметрична относительно главной диагонали, безразлично, возьмем ли мы число из клетки XU или из клетки UX). Затем мы

начинаем выкидывать из графа «лишние» ребра, начиная с того ребра, которому соответствует минимальное число. На каждом следующем шаге снова выкидывается «минимальное» ребро, за исключением тех случаев, когда это делает граф несвязным. Ребра выкидываются до тех пор, пока граф не превратится в дерево, которое и является искомым. Проиллюстрируем принципы работы алгоритма на следующем условном примере (табл. 5 и рис. 7).

Таблица 5

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	...
<i>a</i>		90	35	50	29	...
<i>b</i>	90		80	30	40	...
<i>c</i>	35	80		78	31	...
<i>d</i>	50	30	78		45	...
<i>e</i>	29	40	31	45		...
...

Сначала, по первому основному условию, мы выкидываем ребро ae , являющееся «минимальным» (29), а затем, по тому же правилу, ребра bd (30), ce (31), ac (35), be (40)

в указанной здесь последовательности. На следующем шаге (рис. 7, 2) необходимо выкинуть ребро *ad* (50), а не *de* (45), хотя 50 больше 45, так как удаление ребра *de* изолирует вершину *e* и делает граф несвязным. Ребра, нарисованные сплошной линией, и образуют искомое дерево предложения (см. рис. 7,2)

Аналогичным образом устанавливается дерево синтаксических связей для всех предложений данного текста, и работа алгоритма заканчивается.

Информация о непосредственных синтаксических связях словоформ, при условии что уже проведен морфологический анализ текста, позволяет получить весьма значительную информацию о лексическом аспекте языка и углубить наше понимание его синтаксиса. И. А. Мельчук, например, указал для этих условий способ выделения так

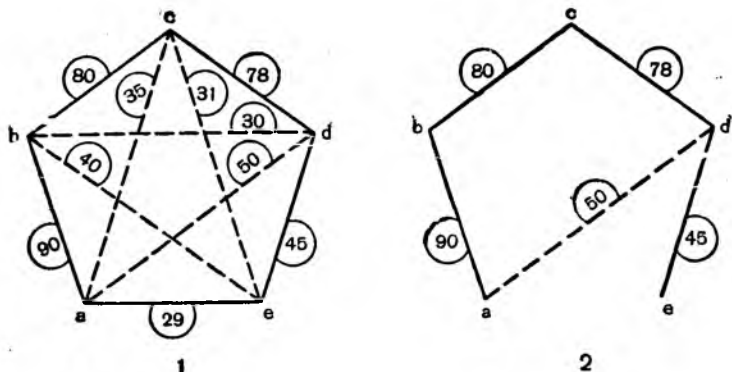


Рис. 7.

называемых устойчивых словосочетаний <112>, а Б. В. Сухотин предложил весьма простой метод переработки информации о непосредственных синтаксических связях словоформ в информацию о зависимостях, или подчинении (метод ориентации графа предложения), а также методы переработки информации о зависимостях в информацию о качественно различных видах синтаксических зависимостей (согласовании, управлении, примыкании). Более того, не выходя за пределы информации, которую дает нам изложенный выше алгоритм Б. В. Сухотина, мы можем получить интересные сведения о значениях слов того или иного языка. На последнем вопросе мы остановим-

ся несколько подробнее и обсудим некоторые результаты выполненного нами описания семантики русского глагола по его синтаксическим свойствам¹.

Предшествующий текст этой книги создает, как нам кажется, достаточное представление о логической структуре моделей интересующего нас класса и технике работы с ними. Поэтому при рассмотрении названной выше темы целесообразно сосредоточиться на содержательной стороне дела. Что касается формальной стороны, то более или менее подробно будет охарактеризован лишь один алгоритм, а из числа формальных понятий будут упомянуты только самые необходимые. К ним относятся понятия управления и совместимости управляемых форм.

Понятие управления может быть выработано алгоритмом, использующим в качестве исходной информацию о словах данного языка, синтаксических зависимостях между словами в составе предложения, а также об основах и грамматических формах слов. Тогда под управлением можно понимать, как это иногда и делается, свойство основы главного слова, обычно глагола, предсказывать форму (или формы) зависимого слова (или слов), ср. *бояться чего, управлять чем, заботиться о ком, дарить кому что, требовать чего от кого, обеспечивать кого чем, вменять что кому во что*.

В принципе каждый глагол способен управлять более чем одной формой, но разными формами он управляет с различной силой. Силу управления можно определить вполне строго, понимая под ней условную вероятность того, что в тексте появится данная форма или совокупность форм при условии, что уже появился данный глагол <74>, <7>. На основе различного рода формальных критериев <7> в множестве форм, управляемых данным глаголом, можно выделить подмножество достаточно сильно управляемых форм. Последние входят в качестве существенной части в число синтаксических свойств глагола.

Между синтаксическими свойствами глагола и его семантическими признаками существует глубокая связь.

¹ Вообще говоря, построенная нами модель относится к классу экспериментальных, а не дешифровочных. Однако в определенных рамках и для определенного материала она сводится к дешифровочной модели (ср. <8>). Интересная попытка описания семантики слов на основе их грамматических и сочетаемостных свойств сделана в работе А. Я. Шайкевича <203>.

Она проявляется в том, что сходству в синтаксических свойствах глаголов соответствует, как правило, сходство их семантических признаков, а различию в синтаксических свойствах — различие семантических признаков. Эта связь возникает неизменно во всяком нормализованном языке, избегающем дублетности форм и стремящемся использовать свои средства последовательно и единообразно, в результате действия аналогии, которая приводит к выравниванию синтаксических свойств слов, семантически близких друг другу. Перестройкой по аналогии объясняется, например, современный тип управления глаголов *наблюдать* (за кем-либо), *руководить* (кем-либо), *трепетать* (перед кем-либо), которые в XIX и даже начале XX века способны были управлять формами над N_i , N_a , N_g соответственно; ср. *наблюдать над разными работами* (Акс.), *руководить читателей* (Л. Толстой), *трепетать мужа* (Чехов). Перестройка была вызвана давлением безусловно господствовавших типов *глядеть* (*надзирать, следить, смотреть*) за кем-либо, *командовать* (*править, распоряжаться, управлять*) кем-либо, *дрожать* (*трястись*) перед кем-либо. Именно благодаря тому, что синтаксические средства языка используются последовательно и единообразно, мы можем безошибочно определить семантический класс глаголов, способных управлять формами *кого чем по чему* (*бить, колотить, ударять*), *кому что о чем* (*говорить, рассказывать*), *кому от кого за что* (*влетело, досталось, попало*) и т. д.

Суждения о сходствах и различиях в значениях глаголов, основанные на сходствах и различиях в их синтаксических свойствах, могут касаться либо множества употреблений одного глагола, либо множества употреблений всех глаголов изучаемого языка. В первом случае речь идет о разграничении значений данного глагола, а во втором — о выделении семантических классов глагольных значений (в дальнейшем для краткости глагольные значения называются глаголами). Обе эти задачи решаются одинаковыми средствами; меняется только материал, к которому последние применяются.

Главным средством установления семантических сходств и различий в некотором множестве употреблений глагола является принцип *с о в м е с т и м о с т и*, который можно сформулировать следующим образом: две управляемые формы совместимы при данном глаголе, если имеется хотя

бы одна фраза с этим глаголом, в которой они реализуются одновременно. В силу этого определения формы N_a , на N_a , N_i при глаголе *отвечать* совместимы, ср. *Он ответил другу на предложение*, *Он ответил другу согласием*, *Он ответил на предложение согласием* и даже *Он ответил другу на предложение согласием*.

Совместимость управляемых форм при данном глаголе свидетельствует о тождестве его значения, а несовместимость форм является показателем семантического различия; исключения из этого правила немногочисленны и могут быть описаны формально. Указанный принцип позволяет разграничить значения глагола *отвечать* во фразах *Он отвечает другу* и *Работа отвечает всем требованиям*: во второй фразе форма дательного падежа несовместима с формами на N_a N_i . Аналогичным образом разграничиваются значения глаголов *бить*, *биться*, *гладить*, *мешать*, *открываться*, *разводить*, *рисоваться*, *трясти* и очень многих других, ср. *бить лошадь кнутом по крупу* и *бить зорю*, *биться головой о стену*, *биться с братом на кулаках* и *биться над задачей*, *гладить дочери лоб ладонью* и *гладить белье утюгом*, *мешать воду с вином* и *мешать кому-либо работать (в работе)*, *открываться другу (перед другом)* во всем и *открываться на улицу* (о двери), *разводить мужа с женой* и *разводить руками*, *рисоваться перед всеми своей выдержкой*, *рисоваться кому-либо в воображении* и *рисоваться на бледном небосклоне* (о силуэте), *трясти мальчика рукой за плечо* и *трясти головой*.

Расширим теперь область применения принципа совместимости, перейдя от отдельного слова к словарю в целом и рассматривая управляемые формы, совместимые хотя бы при одном глаголе, в качестве дифференциальных синтаксических признаков, различающих классы глаголов. Глаголы, имеющие одни и те же синтаксические признаки, объединяются в один класс. Таким образом могут быть получены классы *бить* (*колотить*, *молотить*, *ударять*) кого чем по чему, *биться* (*колотиться*, *стучаться*, *ударяться*) чем обо что, *гладить* (*трепать*) кому что чем, *мешать* (*помогать*) кому *работать (в работе)*, *мешать* (*путать*, *разводить*, *разлучать*, *соединять*, *сравнивать*) что с чем, *открываться* (*виниться*, *исповедоваться*, *каяться*, *признаваться*, *сознаваться*) кому (*перед кем*) в чем, *рисоваться* (*гордиться*, *кичиться*, *хвастаться*, *щеголять*) чем перед кем, *трясти* (*дергать*, *трогать*,

тормозить, хватать) кого чем за что и другие, семантическая однородность которых настолько очевидна, что отпадает необходимость в каком-либо содержательном комментарии.

Совокупность синтаксических признаков определенного класса глаголов будем считать их синтаксической парадигмой (примеры см. в предыдущем абзаце). Выделение в данном языке множества синтаксических парадигм преследует те же цели и имеет тот же эффект, что и выделение парадигм словоизменительных: оно позволяет в простой и ясной форме описать существенные синтаксические свойства различных слов.

Выше мы говорили о том, что в один класс нашей классификации должны попасть глаголы с одними и теми же синтаксическими признаками. Однако число глаголов с абсолютно совпадающими признаками не слишком велико, и если бы мы при построении классификации руководствовались буквальным смыслом нашего правила, классы во многих случаях оказались бы одноэлементными. В действительности для выделения некоторого множества глаголов в качестве класса достаточно не абсолютного совпадения, а относительно большого сходства их синтаксических признаков. Однако это более слабое условие ставит перед нами гораздо более серьезную задачу: оперировать критерием абсолютного совпадения гораздо легче, чем критерием относительного сходства. Для того чтобы судить о последнем, необходимо научиться измерять его степень. Представления о сходстве глаголов могут быть заменены представлениями о каком-нибудь близком ему свойстве, например о расстояниях между глаголами в некотором семантическом пространстве. Тогда в один класс должны попасть глаголы, находящиеся на относительно небольших расстояниях друг от друга (они-то и окажутся максимально сходными по значению).

Стоящую перед нами проблему можно свести, таким образом, к математической задаче, допускающей вполне строгое решение. Необходимо: 1) ввести метрику на множестве глаголов, то есть измерить расстояния между ними (для этого можно воспользоваться информацией об их синтаксических признаках), 2) предложить процедуру построения классов глаголов по таблице расстояний.

Алгоритм решения указанной задачи естественным образом распадается на две части.

В первой части информация о глагольном управлении, выработанная описанными выше средствами, преобразуется в информацию о расстояниях между глаголами; во второй части информация о расстояниях между глаголами преобразуется в информацию о классах глаголов.

Расстоянием (метрическим) называется в математике неотрицательное число, которое ставится в соответствие паре элементов данного множества и обладает следующими свойствами: (1) расстояние элемента до самого себя равно нулю: $\rho(a, a) = 0$; (2) расстояние от элемента a до элемента b равно расстоянию от элемента b до элемента a : $\rho(a, b) = \rho(b, a)$; (3) расстояние от a до b плюс расстояние от b до c больше или равно расстоянию от a до c : $\rho(a, b) + \rho(b, c) \geq \rho(a, c)$. Если такое число может быть поставлено в соответствие любой паре элементов данного множества, то множество называется метрическим пространством.

Сведения о расстояниях между элементами некоторого множества можно легко получить, если имеется какая-либо система из n координат (признаков), на которых каждый элемент имеет определенное значение. Если обозначить числовое значение элемента a на i -й координате через $f_i(a)$, то расстояние между двумя произвольными элементами a и b можно вычислить по формуле:

$$\rho(a, b) = \sum_{i=1}^n |f_i(a) - f_i(b)|,$$

где \sum — знак суммы, а прямые скобки обозначают модуль, т. е. число, взятое по абсолютной величине (без знака). Эту формулу можно прочесть следующим образом: расстояние между двумя элементами равно сумме модулей разностей числовых значений этих элементов по всем координатам, начиная с первой ($i = 1$) и кончая n -й (см. букву над знаком суммы).

Перейдем теперь к нашей задаче. Как легко догадаться, элементами нашего множества являются глаголы, а координатами — управляемые формы. Каждый глагол на каждой координате имеет то или иное числовое значение — вероятность того, что в тексте встретится данная форма при условии, что уже появился данный глагол. Представим эту ситуацию в виде таблицы (данные для нее были действительно получены по текстам, но имеют только

Таблица 6

	N_g	N_a	с N_i	о N_p	N_g от N_g	N_a для N_g	N_a из N_g	N_g в N_p
1. беседовать	0	0	0,82	0,18	0	0	0	0
2. бояться	0,29	0	0	0	0	0	0	0
3. ждать	0,24	0,33	0	0	0,02	0	0	0
4. изменить	0	0,86	0	0	0	0	0	0,14
5. испугаться	0,10	0	0	0	0	0	0	0
6. наладить	0	0,80	0	0	0	0	0	0,10
7. ожидать	0,27	0,48	0	0	0,03	0	0	0
8. построить	0	0,88	0	0	0	0,07	0	0
9. производить	0	0,88	0	0	0	0,04	0	0
10. разговаривать	0	0	0,54	0,08	0	0	0	0
11. создавать	0	0,90	0	0	0	0	0,03	0

иллюстративный смысл: более полное и корректное обсуждение проблемы читатель найдет в <8>).

Пользуясь приведенной выше формулой, эту таблицу можно преобразовать в таблицу расстояний (метрическое пространство) (см. стр. 143).

Клетки ниже главной диагонали не заполняются, так как таблица симметрична. Для того чтобы читатель вполне освоился с несложной математической техникой вычисления расстояний по данным типа тех, которые представлены

Таблица 7

	бесе- довать	бояться	ждать	изменить	испу- гаться	наладить	ожидать	постро- ить	произво- дить	разгова- ривать	создавать
<i>беседовать</i>	0	1,29	1,62	2,00	1,10	1,90	1,78	1,95	1,92	0,38	1,93
<i>бояться</i>		0	0,43	1,29	0,19	1,19	0,53	1,24	1,21	0,91	1,22
<i>ждать</i>			0	0,90	0,52	0,80	0,16	0,85	0,82	1,24	0,83
<i>изменить</i>				0	1,10	0,10	0,82	0,23	0,20	1,62	0,21
<i>испугаться</i>					0	1,00	0,68	1,05	1,02	0,72	1,03
<i>наладить</i>						0	0,72	0,25	0,22	1,52	0,23
<i>ожидать</i>							0	0,77	0,74	1,40	0,75
<i>построить</i>								0	0,03	1,57	0,12
<i>производить</i>									0	1,54	0,09
<i>разговаривать</i>										0	1,55
<i>создавать</i>											0

в таблице 6, вычислим здесь расстояние между глаголами *беседовать* и *бояться*:

$$\rho(\text{беседовать}, \text{бояться}) = \sum_{i=1}^8 |f_i(\text{беседовать}) - f_i(\text{бояться})| = |0 - 0,29| + |0 - 0| + |0,82 - 0| + |0,18 - 0| + |0 - 0| + |0 - 0| + |0 - 0| + |0 - 0| = 0,29 + 0,82 + 0,18 = 1,29.$$

Составлением таблицы расстояний заканчивается первая часть работы алгоритма. Во второй части по таблице расстояний строятся классы, причем формализуется следующая содержательная гипотеза: классы — это множества точек, лежащих недалеко друг от друга (образующих уплотнения в метрическом пространстве). Алгоритм работает с соблюдением следующих условий:

(1) Для любых трех элементов a , b и c , если a и b принадлежат классу K_i , а c не принадлежит ему, то $\rho(a, b)$ должно быть меньше $\rho(a, c)$. Можно доказать, что если на множестве соблюдено это условие, то на нем задано разбиение на попарно непересекающиеся классы (кроме тех случаев, когда один класс целиком включен в другой).

(2) Число классов в разбиении должно быть минимальным, но не меньше двух. Можно доказать, что если на множестве соблюдено и второе условие, то разбиение является единственным (не существует никакого другого разбиения, обладающего обоими указанными свойствами). Это разбиение и является искомым.

Ниже дается содержательное изложение алгоритма.

1. На первом шаге из таблицы выбираются два наиболее удаленных друг от друга элемента. По нашим условиям они не могут относиться к одному классу, и мы получаем два разных класса: {беседовать, ...}, {изменить, ...}.

2. Будем добавлять в них по одному элементу, каждый раз выбирая из таблицы тот глагол, который находится на минимальном расстоянии от какого-нибудь уже выделенного элемента одного из формирующихся классов. Включая новый элемент в некоторый класс, мы должны смотреть, не нарушает ли это первого условия. Если оно не нарушено, то мы повторяем вторую команду. Если первое условие нарушено для какого-нибудь элемента формирующегося класса, то элемент-нарушитель извлекается из него, и образуется новый класс. По этим правилам мы добавим во второй класс *наладить*, *производить*, *создавать*, *построить*, а затем в первый — *разговаривать* (в указанном здесь порядке). Попытаемся теперь включить в класс {беседовать, разговаривать} глагол *испугаться*, так как именно этот еще не расклассифицированный глагол находится на минимальном (0,72) расстоянии от элемента (глагола *разговаривать*), уже попавшего в какой-то класс. Мы обнаруживаем, что этого сделать нельзя, так

как p (*испугаться, беседовать*) = $1,10 > p$ (*испугаться, производить*) = $1,02$. Поскольку первое условие нарушено, глагол *испугаться* должен быть отнесен к новому классу {*испугаться, ...*}.

3. В случае, если образование нового класса не вызывает нарушения первого условия ни для одного из уже выделенных элементов, повторяется команда 2 (в нашем случае дело обстоит именно таким образом, и в третий класс попадают все остающиеся глаголы, т. е. *испугаться, бояться, ждать, ожидать*). Если нарушение имеет место, то все элементы-нарушители извлекаются по одному из уже построенных классов и возвращаются в число нерасклассифицированных элементов. Это продолжается до тех пор, пока нарушение не будет ликвидировано, после чего повторяется команда 2.

Изложенный нами алгоритм дает повод для двух замечаний. Во-первых, достойным внимания кажется нам то обстоятельство, что алгоритм действительно отыскивает искомое разбиение, то есть разбиение множества глаголов на классы, удовлетворяющее первому и второму условиям. Укажем, что это свойство может быть формально доказано. Во-вторых, получающиеся классы семантически достаточно однородны; в частности, три класса, построенные в нашем примере, этим свойством, конечно, обладают, ср. 1) *беседовать, разговаривать*; 2) *изменить, наладить, производить, создавать, построить*; 3) *испугаться, бояться, ждать, ожидать*. Заметим, что синтаксические признаки глаголов двух последних классов не совпадают не только в количественном отношении, но и качественно. Поэтому мы могли бы получить еще более тонкое и адекватное разбиение множества на классы повторным применением алгоритма к материалу второго и третьего классов: 2а) *изменить, наладить*; 2б) *производить, создавать, построить*; 3а) *испугаться, бояться*; 3б) *ждать, ожидать*.

Проведенные автором на более широком материале эксперименты <8> дали неплохие результаты; поэтому можно надеяться на получение семантически содержательных классификаций по синтаксическим данным. Если при этом имеется некоторое количество слов, которым мы можем поставить в соответствие предметы и явления реальной действительности, то в принципе становится воз-

можным не только устанавливать тождества и различия, но и приписывать словам вполне определенные семантические признаки.

Чтобы показать, каким образом можно сделать этот последний шаг, рассмотрим цитированную в другой связи (стр. 103) классическую фразу Л. В. Щербы *Глокая куздра штеко будланула бокра и кудрячит бокрёнка*. Обычно она приводится как пример того, что формальный грамматический разбор предложения возможен без знания лексических значений входящих в него слов. Нам кажется, что анализ этой фразы дает основания для более сильного утверждения: зная язык, на котором она написана, мы можем не только дать ее грамматический разбор, но и установить, в первом приближении, значения входящих в нее слов.

1) *Будлануть* — глагол, обозначающий **на си л ь с т в е н н о е и э н е р г и ч н о е** воздействие на объект, нечто вроде *ударить*; действительно, этот глагол является переходным, поскольку у него есть выраженное одушевленным существительным (ср. окончание *-а* в винительном падеже) прямое дополнение (*бокра*), а переходные глаголы однократного подвида на *-ануть* имеют в русском языке именно это значение, ср. *давануть, долбануть, звездануть, мазануть, резануть, рубануть, садануть, стегануть, толкануть, трепануть, тряхануть, хлестануть, шибануть, щелкануть, щипануть* (ср. В. В. Виноградов <30, 529—530>). Исключение составляет глагол *сказануть*, но значение «сказать» в данном случае невозможно, поскольку в роли прямого дополнения здесь выступает одушевленное существительное (можно *сказать что-то*, но не *кого-то*).

2) *Кудрячить* — глагол с аналогичным значением насильственного воздействия на объект или антонимичным значением типа «ласкать» (ср. <152, 67>); основанием для этого вывода служат следующие факты: *кудрячить* и *будлануть* обладают идентичными синтаксическими свойствами: оба они — переходные глаголы, имеющие в качестве прямого дополнения одушевленное существительное. В данном предложении они являются, кроме того, однородными членами, связанными соединительным союзом *и*. Для однородных членов, имеющих одинаковые синтаксические свойства и связанных союзом *и* (именно *и*, а не *но*, *а* и т. п.), характерно наличие общих семантических признаков (ср. ненормативность фраз типа *Он тряханул меня*

и видит моего брата или Он потрянул меня и благодарит моего брата, сказуемые которых не имеют, должно быть, общих семантических признаков, кроме признака переходности).

3) *Штеко* — наречие, в значение которого входит признак интенсивности (нечто вроде *крепко, как следует*, хотя возможно и полярное по интенсивности значение *слегка*); основанием для этого вывода служит, во-первых, тот факт, что оно образовано от прилагательного *штекий* (ср. *дикий* — *дико*, *крепкий* — *крепко*, *шибкий* — *шибко* и т. п.) и в качестве такового не может быть обстоятельством места, времени, цели, причины и т. п.; во-вторых, тот факт, что оно модифицирует глагол со значением интенсивного насильственного воздействия на объект, а качественные наречия, относящиеся к глаголам с таким значением, не могут не выражать признака интенсивности действия [ср. ненормативность фраз типа *Возница прекрасно (аналогично) хлестанул лошадь*].

4) *Бокр* — животное, самец; *бокрёнок* — его детеныш. Основанием для этого вывода служит то обстоятельство, что *бокр* — одушевленное существительное мужского рода, а *бокрёнок* — одушевленное существительное мужского рода, содержащее тот же самый корень и имеющее в своем составе суффикс *-ёнок*. Такая совокупность формальных признаков характерна в русском языке для пар слов, первое из которых обозначает самца взрослого животного, а второе — его детеныша, ср. *бобр* — *бобрёнок*, *голубь* — *голубёнок*, *жеребец* — *жеребёнок*, *зверь* — *зверёнок*, *козел* — *козлёнок*, *кот* — *котёнок*, *лев* — *львёнок*, *олень* — *оленьёнок*, *слон* — *слонёнок*, *сом* — *сомёнок*, *тигр* — *тигрёнок*, *тюлень* — *тюленёнок*, *угорь* — *угрёнок*.

5) *Куздра* — живое существо, так как только живое существо способно осуществлять такую целенаправленную деятельность, как *будлание*.

Этот анализ объясняет, почему подавляющему большинству не искушенных в лингвистике носителей русского языка, к которым автор обращался с просьбой дать толкование щербовской фразе, представлялась приблизительно одна и та же картина: самка сильно ударила какого-то самца и наносит удары его детенышу (см. также <187, 316 >).

Мы обозрели различные дешифровочные модели, начиная с буквенных и кончая семантическими. Бросим теперь общий взгляд на модели этого типа. Когда мы рассматриваем «выход» некоторых алгоритмов, нам нелегко избавиться от впечатления известной тривиальности получаемых с их помощью результатов; нам кажется, что мы и без того знаем, какие буквы относятся к числу гласных, а какие — к числу согласных; знаем, какие словоформы в предложениях синтаксически связаны, а какие — нет; знаем, наконец, что *беседовать* семантически ближе к *разговаривать*, чем к *испугаться* или *изменить* и т. д.

Мы, однако, не склонны думать, что такой способ оценки модели является наилучшим. Разумнее оценивать модель рассматриваемого здесь типа не по абсолютному богатству результатов, а по богатству результатов относительно объема исходной информации. Если подходить к дешифровочным моделям с этой точки зрения, то нельзя не признать, что достойно удивления не столько то, что результаты бывают бедны или небезупречны, сколько то, что на основе такой скудной исходной информации удастся получить так много верных сведений о языке и надежно обосновать интуитивно выработанные знания.

Высказывались и другие сомнения по поводу моделей исследования первого типа (Ч. Хоккет <310, 45—46>, М. Холлидей <287, 281>, Н. Хомский <199, 466>). Существо этих сомнений состоит в убеждении, что единицы более высоких уровней не строятся непосредственно из единиц предшествующих уровней. Так, важным компонентом морфемы является значение, не свойственное фонемам; значение предложения не складывается непосредственно из значений входящих в него морфем, но включает в себя и значение синтаксической конструкции, лежащей в основе данного предложения. Это подтверждается тем фактом, что предложения, имеющие один и тот же морфологический состав, могут осмысляться различно в зависимости от того, с какой синтаксической конструкцией они ассоциируются, ср. классический латинский пример *amor patris* (родительный субъекта — «любовь отца», родительный объекта — «любовь к отцу»).

Мы не будем подробно разбирать здесь эти сомнения, быть может, отчасти оправданные. Укажем только, что

многие трудности, кажущиеся непреодолимыми в рамках дешифровочных моделей, решаются более гибкими экспериментальными моделями, которые могут рассматриваться как вариант дешифровочных при бесконечно растущем тексте. Обсуждению экспериментальных моделей мы и посвятим следующую главу.

Глава 2

ЭКСПЕРИМЕНТАЛЬНЫЕ МОДЕЛИ

Экспериментальные исследовательские модели преследуют ту же цель, что и разобранные выше дешифровочные: в них тоже ставится задача перехода от текста к «системе», т. е. задача получения по текстовым данным сведений об элементарных единицах языка, классах элементарных единиц, законах сочетания элементов различных классов и т. п. Их отличие от дешифровочных моделей следует искать не в характере вырабатываемой информации, а в методе ее получения и в характере исходной информации. Если в дешифровочных моделях лингвист ограничивает себя чистым наблюдением, фиксируя поведение интересующего его объекта (фонемы, морфемы, конструкции) в естественных условиях текста и подвергая данные наблюдений только математической обработке, то в исследовательских моделях рассматриваемого здесь типа он наблюдает материал и в искусственных условиях, подвергая его всякого рода преобразованиям. Иными словами, в моделях второго типа он использует эксперимент, невозможный в рамках дешифровочных моделей. Для того чтобы иметь возможность проводить эксперименты, лингвисту недостаточно иметь в своем распоряжении текст, потому что всякий реальный текст имеет ограниченную протяженность и, следовательно, не содержит всех принципиально мыслимых в данном языке фраз. В дополнение к тексту лингвист должен располагать множеством правильных фраз данного языка, в принципе бесконечным. Иными словами, для построения экспериментальной модели лингвист должен иметь возможность обратиться к информанту, который по поводу каждой предъявленной ему фразы должен говорить, принадлежит ли

она к числу фраз на его языке или нет (можно так сказать или нет).

Итак, экспериментальные исследовательские модели ставят перед нами два новых вопроса: вопрос о лингвистических экспериментах и вопрос о множестве правильных фраз. На обоих этих вопросах необходимо коротко остановиться, прежде чем мы перейдем к рассмотрению самих экспериментальных моделей.

Лингвистические эксперименты рассматриваемого здесь типа отличаются от решающих экспериментов, рассмотренных в главе I II части (стр. 91—98). Назначение решающих экспериментов состоит в том, чтобы проверить предсказания модели (теории) и таким образом подтвердить или опровергнуть ее. Что же касается рассматриваемых здесь экспериментов, то они являются подсобным средством обработки сырого материала и в этом отношении подобны той технике управляемого опыта, которой пользуется естествоиспытатель, варьирующий одно из условий обстановки при неизменности всех остальных и фиксирующий реакцию объекта на это изменение.

Ниже мы опишем основные экспериментальные приемы, применяемые в современной структурной лингвистике для решения указанных выше задач. Многие из них были известны и традиционной грамматике, хотя и не использовались ею систематически.

1. Д о б а в л е н и е э л е м е н т о в к д а н н о й ф о р м е. а) В русском языке (как, впрочем, и в других языках) из-за активного процесса перехода предложных групп с существительным в наречие довольно трудно решить вопрос о том, являются ли формы типа *сразу* наречиями или сочетаниями предлога с существительным. Л. В. Щерба предлагал для решения этого вопроса следующий эксперимент: вставить между первым и вторым элементом формы прилагательное и посмотреть, какая фраза получается в результате. Если получается неправильная фраза (ср. *с первого разу*) или фраза с сильным изменением смысла (ср. *за границей — за нашей границей*), то перед нами наречие, а в противном случае — сочетание предлога с существительным <224,72>. б) Из-за омонимии предлогов и наречий типа *up, in, on* и многих других в английском языке вопрос о грамматической природе таких

элементов в словосочетаниях типа *to call up the man* — «позвать мужчину» и *to go up the hill* — «идти в гору» без объективных критериев решить довольно трудно. Ч. Фриз <274, 75> предложил для решения этой трудности следующий экспериментальный тест: поставить между глаголом и спорным элементом качественное наречие и проверить, получается ли в результате правильная фраза. Правильность фразы диагностирует предлог (ср. *to go quickly up the hill* — «быстро идти в гору»), а неправильность — наречие (ср. **to call vigorously up the man*). в) Известно, насколько неясны в большом числе языков границы между дополнениями и обстоятельствами. Р. И. Аванесов <1>, а вслед за ним М. В. Панов <137> предложили различать эти члены предложения на основе следующего экспериментального критерия: добавить к спорному элементу с помощью сочинительного союза типа *и*, *но* несомненное обстоятельство и проверить, получается ли при этом правильная фраза. Если получается правильная фраза, то спорный элемент является обстоятельством (ср. *говорить без ошибок* → *говорить без ошибок, но медленно; произносить в нос* → *произносить в нос и нараспев*); в противном случае спорный элемент считается дополнением.

2. О п у щ е н и е э л е м е н т о в и з д а н н о й ф о р м ы. Для иллюстрации существа этого приема сошлемся на следующее рассуждение О. Есперсена: прямое дополнение теснее связано с глаголом, чем косвенное (при наличии у глагола и того и другого), несмотря на тот в н е ш н и й ф а к т, что косвенное дополнение во многих языках помещается ближе к глаголу, чем прямое. В подтверждение своей мысли О. Есперсен приводит тот экспериментальный факт, что косвенное дополнение можно опустить, а прямое — нет, ср. *They offered the butler a reward* — «они предложили лакею вознаграждение» → *→ They offered a reward* — «они предложили вознаграждение», но не **They offered the butler* — «они предложили лакею»; ср. также *A reward was offered* — «была предложена награда», но не **The butler was offered* — «лакею было предложено» <320>.

3. С у б с т и т у ц и я (з а м е н а) э л е м е н т а д р у г и м э л е м е н т о м. По мнению С. Карцевского <324, 109>, различие между так называемыми моторно-некратными глаголами (*бежать, брести, везти, вести, катиться, лезть, лететь, плыть, ползти* и т. п.) и мотор-

но-кратными глаголами (*бегать, бродить, возить, водить, кататься, лазать, летать, плавать, ползать* и т.п.) объективно проявляется в том факте, что во фразах типа *Поезд ползет, как черепаха, Поезд летит стрелой* невозможна замена моторно-некратных глаголов моторно-кратными (ср. ненормативность фраз типа * *Поезд ползает, как черепаха* и * *Поезд летает стрелой*). С. Карцевский, следовательно, предполагает, что в классе глаголов движения можно выделить подкласс моторно-некратных глаголов на том основании, что они, в отличие от моторно-кратных глаголов, могут замещать друг друга в контекстах (окружениях) N_n^1 — как N_n^2 и N_n^1 — N_{comp}^2 (творительный сравнения), которые являются для них «диагностическими», ср.

он	$\left\{ \begin{array}{l} \text{бежит} \\ \text{бредет} \\ \text{едет} \\ \text{катится} \\ \text{летит} \\ \text{плывет} \\ \text{ползет} \\ \text{.....} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{как черепаха} \\ \text{стрелой} \end{array} \right\}$
----	--	---

4. Перестановка элементов данной формы. Для иллюстрации этого случая сошлемся на работу А. М. Пешковского о сочинении и подчинении предложений. Полемизируя с М. Н. Петерсоном, утверждавшим, что различие между сочинением и подчинением «не опирается на внешние данные языков» (143, 109), А. М. Пешковский показывает различие между этими видами связи с помощью следующего несложного эксперимента: возможность перестановки двух простых предложений в составе сложного указывает на наличие связи сочинения между ними, а невозможность такой перестановки свидетельствует о подчинительном характере этой связи, ср. *Язык мой немеет, и взор мой угас* → *Взор мой угас, и язык мой немеет*; *Им можно, а вам нельзя* → *Вам нельзя, а им можно*; но *Он говорит, что хозяин приехал* —

* *Хозяин приехал, что он говорит*; *Он говорит, точно горохом сыплет* — * *Горохом сыплет, точно он*

говорит. Кроме того, подчиненное предложение может быть поставлено в разные места подчиняющего (ср. *Он говорит, точно горохом сыплет* → *Он, точно горохом сыплет, говорит*), а для сочиненного предложения такие перестановки невозможны (ср. неправильность предложения * *Язык, и взор мой угас, мой немее*). В связи с этим А. М. Пешковский характеризовал подчинительный тип как «несимметричный» и «необратимый», а сочинительный тип — как «симметричный» и «обратимый».

5. Трансформация данной формы в другую. Понятие трансформации является более сложным, чем понятие добавления, опущения, субституции или перестановки, и поэтому будет ниже определено. Однако начнем мы, как и во всех предшествующих случаях, с иллюстраций. а) Одним из первых и весьма эффективных опытов применения экспериментальной методики трансформационного анализа к решению некоторых синтаксических вопросов является работа В. Н. Сидорова и И. С. Ильинской о выражении субъекта и объекта действия в русском языке <161>. Описывая структуру предложения, они используют три ряда понятий: 1) семантические (субъект, объект, орудие); 2) синтаксические (подлежащее, прямое дополнение, косвенное дополнение); 3) морфологические (именительный падеж, винительный падеж, творительный падеж). Между парами понятий нет никакой необходимой связи: объект не обязательно является прямым дополнением (ср. подлежащее пассивной конструкции), а прямое дополнение не обязательно выражается винительным падежом. В частности, сочетание предлога *по* с дательным падежом во фразе *Отец дал детям по груше* семантически и синтаксически тождественно винительному падежу во фразе *Отец дал детям грушу* (как и винительный падеж, *по* + дательный падеж синтаксически является прямым дополнением, а семантически — объектом), а сочетание предлога *по* с дательным падежом во фразе *По груше упало с дерева* семантически и синтаксически тождественно именительному падежу во фразе *Груша упала с дерева* (синтаксически оно является подлежащим, а семантически — субъектом). «Таким образом, сочетание предлога *по* с дательным падежом является формой, служащей для выражения субъекта-объекта. По своему употреблению она соотносительна с именительным и винительным падежами...» <161, 347>. Эксперименталь-

ным инструментом, с помощью которого авторы подтвердили этот существенный вывод и ряд других интересных результатов, служил для них, по существу, трансформационный анализ, а именно факт трансформируемости фраз типа *Отец дал детям грушу* во фразы типа *Отец дал детям по груше* и фраз типа *Груша упала с дерева* во фразы типа *По груше упало с дерева*. Трансформации позволили вскрыть глубокое внутреннее сходство внешне различных предложений. б) Трансформационный метод используется для разграничения предложений и словосочетаний различных типов <68>, <81>, <127>, <139>, <180>, <330>, <9>; словосочетания рассматриваются как объекты, а трансформации — как бинарные признаки этих объектов. Составляется таблица, столбцы которой озаглавлены символами типов словосочетаний, а строки — символами трансформаций; клетки таблицы заполняются единицами и нулями в зависимости от того, допускает ли некоторое словосочетание данную трансформацию или нет. Определением словосочетания является столбец таблицы (вектор-столбец); в один класс попадают словосочетания, векторы которых равны ¹. в) Трансформационный метод используется даже для решения ряда семантических вопросов; в частности, в красивом примере З. Харриса *She made him a good husband because she made him a good wife* <295, 271> — «Она сделала из него хорошего мужа, потому что она была ему хорошей женой» каузативное значение глагола *to make* в первом простом предложении отличается от связочного значения *to make* во втором с помощью следующих трансформаций: *She made him a good husband* → *She made*

Таблица 8

	c_1	c_2	c_3	..	c_n
T_1	1	0	1	..	0
T_2	0	1	0	.	0
...
T_i	0	0	0	..	1

¹ Разумеется, эта методика не ограничена ни трансформациями в качестве преобразований, ни словосочетаниями в качестве объектов.

a good husband of him (но не * *She made a good husband for him*); *She made him a good wife* → *She made a good wife for him* (но не * *She made a good wife of him*).

Дадим теперь более строгое определение трансформации¹. Чтобы построить его, нам потребуется: 1) информация о морфемах данного языка типа той, которую дает описанный в предыдущей главе алгоритм З. Харриса (стр. 127), 2) информация о словоформах² и непосредственных синтаксических связях словоформ типа той, которую дает описанный там же алгоритм Б. В. Сухотина, 3) информация о множестве правильных фраз, которую мы предположили заранее заданной для экспериментальных моделей.

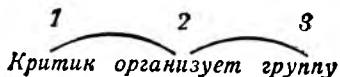
Будем говорить, что две правильные фразы находятся в отношении т р а н с ф о р м и р у е м о с т и, если они имеют (1) одни и те же лексические морфемы и (2) одно и то же дерево непосредственных синтаксических связей. Первое требование исключает из числа трансформируемых многочисленные пары фраз типа *Профессор осматривает пациента* — *Книга стоит рубль*, *Он живет под Москвой* — *Мальчик учится у мастера* и т. п. Правда, оно исключает из числа трансформируемых и некоторое число фраз типа *Москва лежит на восток от Парижа* — *Париж лежит западнее Москвы*, *Он продал машину приятелю* — *Приятель купил машину у него*, которые в определенном смысле разумно было бы считать трансформируемыми, но полезная работа, выполняемая этим требованием, вполне покрывает неизбежные в экспериментальной модели издержки. Но одного этого требования недостаточно; оно не запрещает нам трансформировать фразу *Критик организует группу* во фразы *Организация группирует критиков*, *Организаторы критикуют группу* и т. п., хотя семантически они не имеют ничего общего с исходной фразой³.

¹ Это определение имеет смысл в рамках экспериментальных моделей и отличается от того, которое используется в порождающих грамматиках (см. часть IV).

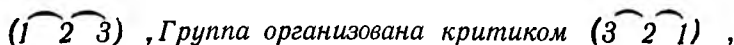
² Под словоформой мы будем здесь понимать лексическую морфему вместе со всеми относящимися к ней грамматическими морфемами, включая связку (обозначаемую через *cop*, от лат. *copula*) и предлог (ср. А.А. Шахматов <213, 432>, А.М. Пешковский <144, 37>, И. И. Мещанинов <119, 296>, Е. Курилович <92, 180>, <89, 66>, отчасти В. В. Виноградов <30, 698>).

³ Ср. пример О. Есперсена <320, 332>: *The writer, when he sits down to commence his novel, should do so, not because he has to tell a*

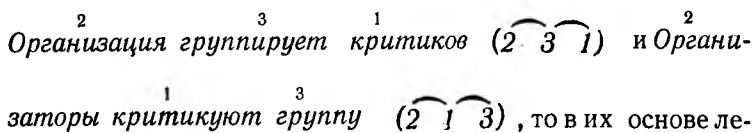
Чтобы запретить такие бессмысленные трансформации, и вводится второе условие. Оно требует, чтобы во фразах были одни и те же отношения непосредственной синтаксической связи между лексическими морфемами. Поясним это на нашем примере. Перенумеруем лексические морфемы фразы *Критик организует группу* и обозначим дугами отношения непосредственной синтаксической связи:



Как легко заметить, в вершине дерева стоит словоформа, содержащая лексическую морфему № 2. То же самое дерево находим во фразах *Критик — организатор группы*



и поэтому любые две фразы этого множества находятся в отношении трансформируемости. Что касается фраз



жат другие синтаксические деревья (с лексическими морфемами № 3 и № 1 в вершинах), и, следовательно, они не находятся в отношении трансформируемости к исходной фразе¹.

Отметим, что если соблюдены два указанных здесь чисто формальных требования, то это автоматически при-

story, but because he has a story to tell—«Когда писатель садится за роман, он должен делать это не потому, что на нем лежит обязанность что-то рассказать, а потому, что ему есть что рассказать».

¹ В основных чертах это определение было сформулировано З. Харрисом (195), (292): другие определения трансформируемости см. в (152), (180), (81), (201), (208), (209), (211). Приведенное в тексте определение упрощено и поэтому обладает пониженной эффективностью, но в принципе оно может быть усилено до нужной степени.

водит к сохранению значительной части первоначального смысла фразы. Очевидно, какую роль играет при этом идентичность лексических морфем, но менее очевидно, каким образом идентичность синтаксических деревьев может способствовать сохранению смысла фразы. Объяснение этому мы находим в том факте, что синтаксис в большей мере, чем какая-либо другая часть грамматики, связан с семантикой; поэтому наличие синтаксических связей между словами является признаком их семантической связанности¹. Требование идентичности лексических морфем — это формализованное условие сохранения смыслов; требование идентичности синтаксических деревьев — это формализованное условие сохранения связей между смыслами.

Таким образом, оставаясь на вполне формальном уровне, мы добиваемся довольно хорошего контроля над значением, а формализация семантики, и в особенности синонимии, является одной из главных целей трансформационного анализа <69>.

Две фразы, находящиеся в отношении трансформируемости, называются *т р а н с ф о р м а м и*. Любые два трансформанта, объединенных знаком трансформируемости (\rightarrow или \leftrightarrow), образуют *т р а н с ф о р м а ц и ю*. Термин «трансформация» используется в экспериментальных моделях еще в одном смысле. Поясним его на примере. Возьмем пару трансформов и заменим каждую словоформу в каждом из них символом класса словоформ, к которому она принадлежит. Символы классов словоформ объединим дугами непосредственной синтаксической связи так, как объединены соответствующие словоформы во фразах; полученные цепочки символов, называемые конструкциями, свяжем стрелкой трансформируемости, ср.

критик организует группу \longrightarrow *группа организована критиком*



Результат, т. е. пара полученных таким образом конструкций (в частности, $N_n^1 V N_a^2 \leftrightarrow N_n^2 \text{cop} A(V) N_i^1$), также назы-

¹ Возможны случаи, когда между словоформами имеется семантическая связь, но нет синтаксической; тогда нам не удастся использовать некоторых «хороших» трансформаций, поскольку они не

вается трансформацией. В данной главе термин «трансформация» будет использоваться именно в этом смысле. Если у нас будет необходимость говорить о трансформациях в первом смысле, мы будем называть их фразовыми трансформациями.

6. П е р е в о д. Во всех рассмотренных выше случаях мы имели дело с преобразованием, или переводом, одной языковой формы в другую внутри одного и того же языка. В качестве экспериментального приема используется и самый обычный «внешний» перевод. В частности, Дж. Гринберг <280> применил этот прием для решения одной из основных задач морфологического анализа — задачи выделения морфов. Дж. Гринберг считает, что сегментация некоторой формы *AB* на два морфа, *A* и *B*, оправдана в том случае, если в данном языке есть «квадрат», т. е. четыре формы вида *AB*, *CB*, *AD* и *CD*, и если существует язык, на котором переводы этих четырех форм также образуют квадрат. С этой точки зрения формы в первом из приводимых ниже примеров морфологически членимы (их переводы, например, на английский язык образуют квадрат), а формы во втором — нет.

- | | |
|-------------------------------------|-------------------------------------|
| (1) <i>добр-ее</i> → <i>kind-er</i> | <i>слеп-ее</i> → <i>blind-er</i> |
| <i>добр-ота</i> → <i>kind-ness</i> | <i>слеп-ота</i> → <i>blind-ness</i> |
| (2) <i>коло-ть</i> → <i>chop</i> | <i>коло-туть</i> → <i>beat</i> |
| <i>моло-ть</i> → <i>grind</i> | <i>моло-туть</i> → <i>thrash</i> |

Заметим, что формы из примера (2) допускают еще несколько разбиений границами (*к-олоть*, *м-олоть*, *к-олотить*, *м-олотить*; *ко-лоть*, *мо-лоть*, *ко-лотить*, *мо-лотить* и т. д.), ни одно из которых экспериментально (переводами) не подтверждается.

Перейдем ко второму поставленному в начале данной главы вопросу — вопросу о м н о ж е с т в е п р а в и л ь н ы х ф р а з. Понятие правильной фразы не имеет общепринятого формального определения и обычно поясняется с помощью содержательных рассуждений следующего типа: понятие правильности относится только к грамматическому оформлению фразы; фраза *Он пришла* непра-

удовлетворяют второму условию трансформируемости, ср. *Книга пленяет нас юмором* — *Юмор книги пленяет нас*. Ошибки такого рода с избытком компенсируются той полезной работой, которую выполняет второе требование.

вильна, а фраза *Бесцветные зеленые идеи бешено спят* правильна, хотя и бессмысленна. Грамматическую правильность не следует поэтому путать с лексической и стилистической правильностью, осмысленностью, истинностью, ложностью и тому подобными категориями <152>, <195>, <197>, <199>, <201>, <350>, <360>, <361>, <382>.

Такие рассуждения, во-первых, не вполне корректны, так как предполагают, что исследователю уже известны все тонкости грамматики данного языка, его лексические и стилистические нормы, и, во-вторых, неэффективны, так как не дают возможности ориентироваться в реальной картине текстовых данных, гораздо более сложной и пестрой, чем можно заключить по тривиальным примерам типа *Груша растет на сосне* или *Он пришла*. Вот некоторые примеры:

(1) *Вот и все о двух ничтожествах..., людях нечистоплотных ни в политическом, ни в моральном отношении* («Известия», 11/IX—64).

(2) *Как тускло все не было кругом!*

(3) *Бозинг навязчиво внушает читателю, что вслед за Гитлером на исход битвы у Волги пагубно сказалась слабость армий союзников Германии* («Известия», 30/I—63).

(4) *...Пожелания будут внимательно рассмотрены, имея в виду оказать... помощь* («Правда», 20/IX—64).

(5) *Накал битвы... передан писателем не столько через впечатляющие батальные сцены, сколько в предельном напряжении духовных сил героев романа* («Известия», 29/V—64).

(6) *О том... пояснять не требуется* («Известия», 25/XI—64).

(7) *Высказать им наше отношение на роль... печати* («Известия», 27/X—63).

(8) *Сержант побывал дома за успехи в учебе* («Правда», 23/II—63).

(9) *Лично за собой никаких аморальных и других поступков поведения на стадионе не проявлял* (Из заявления обвиняемого, «Известия», 28/VIII—64).

(10) *...Правильное сочетание научной молодежи... с мудростью и опытом старшего поколения* («Известия», 10/VII—64).

(11) *Ермолай дремал, сидел около меня и клевал носом* или *Ермолай дремал, клевал носом и сидел около меня.*

Трудно сказать, какие из этих фраз грамматически неправильны, а какие — лексически, семантически или стилистически (ср., например, последнюю фразу—(11)). По-видимому, заранее обречены на неудачу попытки дать формальное определение грамматической правильности фразы для конкретного языка до тех пор, пока не фиксирована вполне точно грамматика данного языка. Поэтому более разумно полагаться на суждение информанта, который должен отвечать на вопрос о том, правильна ли (нормальна ли, принадлежит ли к его языку) данная фраза, а не на вопрос о том, правильна ли данная фраза грамматически. Таким образом, мы действительно принимаем понятие правильности фразы без определения, считая, что множество правильных фраз задано нам в языковом опыте того лица, которое было выбрано нами в качестве информанта.

Посмотрим теперь, каким образом применяются изложенные выше принципы к решению тех или иных лингвистических задач. Из-за обилия имеющихся материалов мы вынуждены будем довольствоваться рассмотрением лишь некоторых вопросов, главным образом из области морфологии, словообразования и синтаксиса. В качестве иллюстраций будут взяты работы, в которых использовался какой-нибудь из описанных выше экспериментальных методов и которые дали, по нашему мнению, нетривиальные результаты.

Мы начнем с м о р ф о л о г и и. Многие вопросы, составляющие содержание обычной морфологии, сводимы к вопросу о к л а с с а х. Части речи со всеми обычно выделяемыми разрядами внутри каждой из них (существительные конкретные и абстрактные, нарицательные и собственные, исчисляемые и неисчисляемые, единичные и собирательные; глаголы переходные и непереходные, предельные и непредельные, действий и состояний и т. д.) представляют собой некоторые классы слов. Любую парадигму можно представить в виде класса словоформ. Наконец, любая грамматическая категория, например категория падежа, числа, рода, времени, вида и др., может быть вполне адекватно задана несколькими большими классами слов или словоформ (категория вида в русском языке — двумя списками глаголов, категория числа —

двумя списками словоформ, категория падежа — шестью списками словоформ <152> и т. д.).

В свое время было предложено использовать для построения таких классов принцип субституции, существо которого состоит в следующем: два элемента входят в один класс, если они способны замещать друг друга в одних и тех же окружениях без нарушения правильности данной фразы (в противном случае они, разумеется, входят в разные классы) <253>, <274>, <289>, <295>, <46>¹. В более точной формулировке, данной в рамках теоретико-множественной модели языка <47>, <88>, <152>, <186>, два элемента, x_i и x_j , считаются эквивалентными (т. е. находятся в отношении субституции и входят в один класс, или «семейство»), если для любой правильной фразы вида $A_1x_iA_2$ существует правильная фраза вида $A_1x_jA_2$, и наоборот, для любой правильной фразы вида $B_1x_jB_2$ существует правильная фраза вида $B_1x_iB_2$.

Очевидно, что в зависимости от того, что считается элементом, окружением элемента и правильной фразой (формой), субституция дает более или менее дробные классы: чем ближе к текстовым единицам элемент и его окружение и чем жестче ограничения на правильность фразы, тем более дробными получаются классы. Так, в теоретико-множественной концепции в качестве элемента рассматривается некоторая словоформа; ее окружением считается остальная часть фразы; наконец, правильными, судя по результатам <152, 66>, считаются не только фразы типа *Идея яростно спит* и *Слон прыгнул на аудиторию*, но и фразы типа *Он выл, когда шакал* (ср. *Он выл, как шакал*); *Плохо, чтобы он пришел* (ср. *Сомнительно, чтобы он пришел*); *Я подошел к столу, которого увидел* (ср. *Я подошел к человеку, которого увидел*). В результате получаются «семейства», которые можно интерпретировать как классы словоформ с идентичным набором грамматических категорий, ср. 1) *столу, окну, человеку, клопу...*; 2) *хохотал, прыгал, бегал, визжал...*; 3) *хорошо, плохо, сомнительно...*; 4) *когда, где, как...* и т. д. В противоположность этому в классической дистрибутивной лингвистике <274>, <295> в качестве элементов рассматриваются основы, их окружением считаются грамматические морфемы или классы основ и грамматических морфем, а на правильность фраз нала-

¹ См. также <34>, <198>, <283> и др.

гаются более жесткие ограничения; в частности, фразы типа *Он выл, когда шакал*; *Плохо, чтобы он пришел* и т. п. правильными не считаются. Иллюстрацией этих принципов является классификация З. Харриса <289>. Ее можно представить в виде процедуры (процесса) переработки языковых фактов, которая на «входе» получает словарь отдельных морфем и основ данного языка, а на «выходе» выдает информацию о классах этих элементов. Процесс получения классификации является ступенчатым. На любом шаге этого процесса вырабатывается какой-то фрагмент конечной классификации (строится некоторый класс элементов), причем на каждом последующем шаге для построения того или иного класса может быть использована информация, полученная на любом из предыдущих шагов, но не может быть использована информация, вырабатываемая позднее. Процесс начинается с выделения основ существительных (класса *N*) в окружении *the -s*, т. е. после определенного артикля и перед окончанием множественного числа. Основы прилагательных и глаголов (*A* и *V*) выделяются в окружениях, состоящих из определенных грамматических морфем и символа *N*, а для выделения основ наречий (*D*) используются окружения с символом *A*. В результате получают классы основ с подклассами внутри каждого из них, классы служебных элементов и т. д. Процесс, который можно уподобить «цепной реакции», охватывает все новые пласты материала, пока классификация не будет завершена ¹.

Богатые и еще не до конца использованные возможности анализа окружения хорошо иллюстрирует одна из работ А. А. Зализняка <62>, получившего или, может быть, подтвердившего с его помощью весьма нетривиальные результаты в области грамматической классификации русских существительных.

В основе работы А. А. Зализняка лежит понятие **с о г л а с о в а т е л ь н о й с в я з и**: это любая синтаксическая связь между существительным и согласуемым словом, при которой форма согласуемого слова зависит от выбора существительного или от формы, в которой оно стоит. Помимо атрибутивной и чисто предикативной связи, к этому типу синтаксической связи относятся: 1) связь

¹ Ср. критические замечания о таких классификациях в <244>, <328>, <330>.

через предлог *из* (ср. *один из домов*, но *одна из стен*), 2) связь через глагол (ср. *эту затею считают безнадежной*, но *это дело считают безнадежным*), 3) связь со словом *который* в придаточном предложении (ср. *дом, в котором*, но *книга, в которой*) и некоторые другие.

Через понятие согласовательной связи определяется ключевое понятие **согласовательного класса**. «Согласовательным классом называется такая совокупность существительных, что любые два ее члена, будучи взяты в любой грамматической форме (но одной и той же для обоих), требуют при любом типе согласовательной связи одной и той же словоформы... любого согласуемого слова» <62>.

Во многих языках, имеющих категорию рода (ср. латынь, немецкий и др.), согласовательные классы совпадают с родами. В русском языке из-за наличия в нем категории одушевленности — неодушевленности это совпадение не имеет места, и поэтому число согласовательных классов минимум вдвое превышает число родов:

Род Одушевленность	I мужской	II женский	III средний
Неодушевленность	класс 1 (дом)	класс 3 (стена)	класс 5 (окно)
Одушевленность	класс 2 (врач)	класс 4 (коза)	класс 6 (чудовище)

Существительное любого из шести согласовательных классов в любой из двенадцати форм можно опознать в некоторых диагностических окружениях. Так, для существительных различных согласовательных классов, стоящих в форме именительного падежа единственного числа, диагностическими окружениями являются контексты:

- I. 1. *Вот тот — , который вы видели.*
2. *Вот тот — , которого вы видели.*
- II. 3. *Эта — была из тех, которые вы видели.*
4. *Эта — была из тех, которых вы видели.*

- III. 5. *Это — было из тех, которые вы видели.*
6. *Это — было из тех, которых вы видели.*

Дадим еще диагностические окружения для существительных различных согласовательных классов, стоящих в форме родительного падежа множественного числа, ср.

- I. 1. *Я видел один из этих —.*
2. *Я видел одного из этих —.*
II. 3. *Вот одна из тех —, которые вы видели.*
4. *Вот одна из тех —, которых вы видели.*
III. 5. *Вот одно из тех —, которые вы видели.*
6. *Вот одно из тех —, которых вы видели*¹.

Классификация существительных по шести согласовательным классам не является исчерпывающей. Действительно, чтобы отнести некоторое существительное к одному из шести классов, мы должны определить его значение по признаку «одушевленность — неодушевленность» и его значение по признаку «род». Классификация по первому признаку не представляет затруднений ни для одного существительного русского языка (см. соответствующие диагностические окружения). Что касается классификации существительных по признаку рода, то она не вызывает затруднений лишь до тех пор, пока мы не подойдем к существительным *pluralia tantum*, у которых либо совсем нет форм единственного числа (ср. *ребята, девочки*), либо форма единственного числа омонимична форме множественного [ср. *сани, брюки*, в контекстах *одни новые сани (брюки)* и *многие новые сани (брюки)*]. С помощью диагностических окружений, используемых А. А. Зализняком, существительные первой группы *pluralia tantum* можно отнести к одному из трех родов, ср. *Я видел одного из этих ребят*, но *Вот одна из тех девочек, которых вы видели*. Поскольку, как показывают приведенные диагностические окружения, существительные этой группы

¹ Из этого, между прочим, следует, что категорию рода и одушевленности некоторого существительного можно определить по любой из его форм, если она стоит (или поставлена) в диагностическом окружении. Таким образом, общепринятое мнение о том, что во множественном числе роды не различаются, а противопоставление по одушевленности — неодушевленности выражается только в винительном падеже, нуждается если не в пересмотре, то во всяком случае в существенном уточнении.

являются одушевленными, они относятся либо ко второму согласовательному классу (*ребята*), либо к четвертому (*девчата*). Что касается существительных второй группы *pluralia tantum* (*сани*), то, хотя они относятся к числу неодушевленных, они не могут быть помещены ни в первый, ни в третий, ни в пятый согласовательный класс, так как не обнаруживают «большой близости ни к одному из трех родов. От всех трех родов сразу они отличаются омонимией словоформ разных чисел» <62>. Поскольку ни к одному из уже имеющихся классов их отнести нельзя, а в корректно построенной классификации каждый элемент должен входить в какой-то класс, нам не остается ничего другого, как образовать особый, седьмой, согласовательный класс и одновременно признать, что признак рода в русском языке принимает не три значения, как было принято думать до сих пор, а четыре: мужской, женский, средний и парный (групповой). Следовательно, адекватной грамматической классификацией существительных в русском языке является следующая классификация:

Род Одушевленность	I мужской	II женский	III средний	IV парный
Неодушевленность	класс 1 (дом)	класс 3 (стена)	класс 5 (окно)	класс 7 (сани)
Одушевленность	класс 2 (врач)	класс 4 (коза)	класс 6 (чудовище)	—

Эта классификация и является основным результатом А. А. Зализняка. Выявленные семь согласовательных классов охватывают все русские существительные; иначе говоря, они образуют действительную классификацию существительных в отличие от традиционной системы родов.

Мы ограничимся этими иллюстрациями применения экспериментальных методов в области морфологии и перейдем к словообразованию. Чтобы изучить механизм словообразования, необходимо прежде всего:

1) Указать в множестве слов данного языка все классы слов, связанных отношением производности, или, что то же самое, для каждой лексической (корневой) морфемы данного языка указать все слова, в которые она входит. Иными словами, необходимо установить такое отношение эквивалентности на множестве слов, которое порождает их разбиение на так называемые «словообразовательные гнезда». Действие различных морфологических процессов, таких, как опрощение и переразложение, делает решение этого вопроса гораздо более трудным, чем кажется на первый взгляд. Неясно, например, входят ли в одно «словообразовательное гнездо» слова *вкус* и *кусать*, *деятель*, *надеяться* и *действовать*, *обращать* и *вращать*, *быть*, *забыть*, *добыть* и *убыток*, *предлагать* и *предложение*, *уравнять* и *уравнение*. Интересно, что В. А. Богородицкий считал слово *забыть* опрощенным <21, 98>, а Г. О. Винокур — производным <317, 315>.

2) После того как найдены все «словообразовательные гнезда», необходимо внутри каждого из них установить направление производности, т. е. для любой пары слов, относящихся к данному «гнезду», указать, существует ли правило образования одного из них от другого, и если да, то которое из них является исходным (содержит производящую основу), а которое — производным. Иными словами, на множестве слов, принадлежащих к одному «словообразовательному гнезду», необходимо установить некоторое отношение (частичного) порядка. Известно, как труден этот вопрос, особенно в случае пар типа *работа* — *работать*, *дело* — *делать*, *игра* — *играть* и т. п.

Помимо этого, для каждой пары слов, для которой уже указано направление производности, необходимо установить, входит ли она в *продуктивный* словообразовательный тип; для каждого словообразовательного типа необходимо указать зону его действия и отношение к другим словообразовательным типам (в частности, синонимическим) и т. д.

Ниже мы займемся только первыми двумя вопросами и проиллюстрируем пути их решения на примере работ З. М. Волоцкой <39> и П. А. Соболевой <170>, в которых используются *трансформационные* *методики*.

Для того чтобы яснее представить себе задачу, решаемую в работе З. М. Волоцкой, введем понятия синтаксической и морфологической производности. Будем говорить, что два слова C_1 и C_2 входят в отношение морфологической производности, если они содержат один и тот же корень (одну и ту же лексическую морфему), ср. *рыба* — *рыбак*, *воск* — *вощить*, *стена* — *стенной*, *земля* — *земляной*, *земля* — *земной*, *вода* — *наводнять*, *решать* — *решительность* и пр. Мы можем установить отношения морфологической производности, если у нас имеется способ обнаружить все разные лексические морфемы данного языка. По существу, это — одна из задач морфологического анализа.

Морфологическая производность является необходимым условием синтаксической производности. Слова C_1 и C_2 входят в отношение синтаксической производности¹, если они находятся в отношении морфологической производности и если фразы, содержащие слово C_1 , находятся в отношении трансформируемости с фразами, содержащими слово C_2 . Таковы пары слов *ненавидеть* и *ненависть* (*он ненавидит лесть* ↔ *его ненависть к лести*), *интересовать* и *интересный* (*книга интересуется его* ↔ *книга интересна ему*) и т. д. Следовательно, понятие морфологической производности шире понятия синтаксической производности: если C_1 и C_2 синтаксически производны друг от друга, они и морфологически производны; обратное неверно.

По существу З. М. Волоцкая предлагает метод установления только синтаксической производности (для пар слов, принадлежащих разным частям речи), а вопрос о морфологической производности предполагается уже решенным; иными словами, предполагается заранее заданной информация о различиях и тождестве лексических морфем. Слова, содержащие одну и ту же лексическую морфему, помещаются в типичные для них фразы (словосочетания), например ***быстрое* движение, *быстро* вращать, *прибавить* число, *возвести* в степень** и т. п.

Задаются списком трансформации, например:

1. $A_n N_n \leftrightarrow N(A)_n N_g$ (*быстрое движение* ↔ *быстрота движения*)
2. $V N_a \leftrightarrow N(V)_n N_g$ (*прибавляю число* ↔ *прибавление числа*)

¹ Ср. понятия лексической и синтаксической деривации у Е. Куриловича <89>.

3. $VN_i \leftrightarrow N(V)_n N_i$ (управляю процессом \leftrightarrow управление процессом)
4. $V \text{ в } N_a \leftrightarrow N(V)_n \text{ в } N_a$ (возвожу в степень \leftrightarrow возведение в степень)
5. $N_n^1 N_g^2 \leftrightarrow A(N^2)_n N_n^1$ (права автора \leftrightarrow авторские права)
6. $N_n^1 \text{ из } N_g^2 \leftrightarrow A(N^2)_n N_n^1$ (предмет из металла \leftrightarrow металлический предмет)

Слова C_1 и C_2 считаются производными друг от друга, если для любого типичного словосочетания, содержащего слово C_1 , в этом списке найдется трансформация, переводящая его в типичное словосочетание со словом C_2 . Таковы пары слов *сложный* — *сложность*, *быстрый* — *быстрота*, *автор* — *авторский* и многие другие. Содержательно этот случай можно интерпретировать как производность во всех значениях. Два слова C_1 и C_2 считаются частично производными, если такие трансформации существуют лишь для некоторых типичных фраз с каким-либо из этих слов. Таковы пары *предлагать* — *предложение*, *управлять* — *управление* и другие: словосочетание *предлагать перемирие* можно преобразовать в словосочетание *предложение перемирия* с помощью второй трансформации, но не существует трансформации, которая переводила бы словосочетание *восклицательное предложение* в типичное и правильное словосочетание с глаголом *предлагать*; точно так же словосочетание *управлять процессом* переводимо в словосочетание *управление процессом*, но для словосочетания *центральное статистическое управление* не существует соответствующего правильного словосочетания с глаголом. Этот случай можно содержательно интерпретировать как производность в некоторых значениях. Наконец, слова C_1 и C_2 считаются непроизводными, если ни для одного типичного словосочетания с первым словом не существует трансформации, переводящей его в типичное словосочетание со вторым словом. Таковы слова *уравнять* и *уравнение*, *пух* и *пушной* (*пух* связано с *пуховой*, а *пушной* — с *пушнина*) и т. д.

Этот метод имеет, конечно, ограниченное значение (с его помощью нельзя, например, решить вопроса о наличии производности в парах *рыба* — *рыбак*, *воск* — *вощить*, *привыкнуть* — *отвыкнуть* и т. п.), но в определенной узкой области он является более эффективным

экспериментальным инструментом, чем предлагавшиеся ранее семантические тесты¹.

Перейдем ко второму вопросу — установлению направления производности внутри уже найденного класса (синтаксически) производных слов.

На первый взгляд вопрос кажется достаточно ясным: производным следует считать то слово (ту основу), которое было фактически образовано от какого-то другого, более простого слова (или более простой основы), а исходным следует считать то слово, от основы которого было фактически образовано какое-то другое слово. Это решение неудачно, так как во всех случаях навязывает нам историческую точку зрения на предмет. Между тем, как было выяснено выше, исторические процессы образования слов далеко не всегда соответствуют складывающимся в синхронном состоянии языка отношениям между словами. Вспомним процессы опрощения, переразложения, обратного словообразования и другие им подобные, существо которых сводится к тому, что слово, являющееся производным с исторической точки зрения (*в-кус, де-л-о, зонт*), оказывается простым с точки зрения синхронных отношений, а исторически простое слово (*зонтик*) превращается под влиянием аналогии в производное. По мнению С. Карцевского <324, 49>, В. В. Виноградова <30, 433>, отчасти А. В. Исаченко <78, 46>, такое же обращение отношений имело место в парах типа *делать — дело, работать — работа, думать — дума, играть — игра* и т. п.: исторически глаголы были произведены от существительных, но в современном языке, под влиянием аналогии <30, 433>, существительные воспринимаются как производные от глаголов. Таким образом, мы вынуждены покинуть историческую почву и искать синхроническое решение.

С синхронической точки зрения кажется естественным решение рассматривать в качестве производного структурно более сложное слово (слово с бóльшим числом словообразовательных аффиксов), а в качестве исходного — структурно более простое слово (слово с меньшим числом

¹ Ср., например, тест Г. О. Винокура: «...Значение слов с производной основой всегда определимо посредством ссылки на значение соответствующей первичной основы...» <37, 421>. Отметим, что по существу трансформационная методика установления производности, разработанная З. М. Волоцкой, является формализацией этого содержательного критерия Г. О. Винокура.

словообразовательных аффиксов). Это решение, однако, обслуживает всего один, правда, наиболее типичный, но тем не менее частный, случай, а именно случай аффиксации. Между тем словопроизводство осуществляется не только с помощью аффиксации, но и с помощью других средств, например чередования звуков¹ (ср. англ. *sing* — «петь» — *song* — «песня»), переноса ударения (ср. англ. *'imprint* — «отпечаток» — *im'print* — «отпечатывать»), чередования звуков и переноса ударения (ср. *гнило́й* — *гниль*, *нечи́стый* — *не́чисть*), конверсии² (ср. англ. *doc-tor* — «доктор» — *to doctor* — «врачевать», нем. *sein* — «быть» — *Sein* — «бытие», фр. *dîner* — «обедать» — *dîner* — «обед»; ср. рус. *печь* — *печь*). Во всех этих случаях безаффиксальной производности решение о направлении производности, очевидно, должно быть более сложным и более интересным.

По нашему мнению, наиболее перспективный путь исследования рассматриваемого здесь вопроса был указан еще в 1933 году О. Есперсеном: «...Чтобы обнаружить,— писал он,— к какому классу принадлежит то или иное слово, недостаточно проанализировать его форму саму по себе; решающим является то, каким образом данное слово «ведет себя» по отношению к другим словам в связанной речи и каким образом другие слова ведут себя по отношению к нему» <320, 71>. По существу именно по этому пути пошла П. А. Соболева, предложившая простую, но исключительно эффективную трансформационную методику установления направления производности в парах английских слов вида «глагол — существительное», связанных отношением конверсии. Хотя ее метод был первоначально предназначен для решения частного вопроса, он допускает довольно интересное обобщение, о котором мы скажем ниже несколько слов.

В качестве исходного материала П. А. Соболева берет пары простых и производных слов вида $V \rightarrow N$ и $N \rightarrow V$,

¹ См. критерий Б. Блока и Дж. Трейджера для этого случая <251, 62>.

² Конверсия определяется либо как «переход одной части речи в другую», либо как «безаффиксальный способ словообразования», либо как «словообразование с помощью нулевого аффикса». Наиболее глубоко нам кажется принадлежащее А. И. Смирницкому определение конверсии как такого способа словообразования, при котором единственным словообразовательным средством является грамматическая (словоизменятельная) парадигма слова <167>.

в которых направление производности выражено явно, т. е. с помощью словообразовательных аффиксов, ср. *elect* — «избирать» → *election* — «избрание» (простой глагол и производное существительное) и *friend* — «друг» → *befriend* — «относиться по дружески» (простое существительное и производный глагол). Как и в работе З. М. Волицкой, простые и производные слова помещаются в типичные для них фразы, а фразы связываются трансформациями. Получающуюся картину можно представить в виде следующей схемы:

Трансформация	$V \rightarrow N$	$N \rightarrow V$
1. $N^1VN^2 \rightarrow N^1copN^3N^2's$	<i>He follows the man</i> → <i>He is the man's follower</i> — „Он следует за (этим) человеком → Он — последователь этого человека“.	<i>J befriended Paul</i> → <i>I am Paul's friend</i> — „Я по-дружески отнесся к Полю → Я друг Поля“
2. $N^1V \rightarrow N^1copPN^2$	<i>He is entering (the house)</i> <i>he is at the entrance (to the house)</i> — „Он входит в дом → Он у входа в дом“.	<i>The entrained</i> → <i>they were in the train</i> — „Они сели в поезд → Они были в поезде“
3. $N^1VN^2 \rightarrow N^2copPN^3N^1's$	<i>Paul exhibited the pictures</i> → <i>the pictures were at Paul's exhibition</i> — „Поль выставил картины → Картины были на выставке Поля“	<i>Paul encaged the bird</i> → <i>the bird was in Paul's cage</i> — „Поль посадил птицу в клетку → Птица была в клетке у Поля“.
4. $N^1VN^2 \rightarrow N^2'sN^3PN^1$	<i>The committee elected John</i> → <i>John's election by the committee</i> — „Комитет избрал Джона → Избрание Джона комитетом“	—
5. $N^1VPN^2 \rightarrow N^1'sN^3PN^2$	<i>John differs from the rest</i> → <i>John's difference from the rest</i> — „Джон отличается от всех → Отличие Джона от всех“	—
6. $N^1VD \rightarrow N^1'sAN^2$	<i>He arrived unexpectedly</i> → <i>his unexpected arrival</i> — „Он приехал внезапно → Его внезапный приезд“	—

Как следует из схемы, первые три трансформации характерны и для пар (направления производности) вида $V \rightarrow N$, и для пар (направления производности) вида $N \rightarrow V$. Три последних трансформации характерны исключительно для пар вида $V \rightarrow N$.

Мы выяснили, таким образом, какие трансформации **д и а г н о с т и р у ю т** (обнаруживают) направление производности $V \rightarrow N$ (4—6). Имея в руках диагностические трансформации, мы переходим к парам существительных и глаголов, связанных конверсией, т. е. парам, в которых направление производности внешне никак не выражено, и делаем основной шаг, которым является **з а к л ю ч е н и е п о а н а л о г и и**: если пары вида «глагол — существительное» с невыраженным направлением производности допускают какие-нибудь диагностические трансформации, то глагол в них является исходным (простым) словом, а существительное — производным; в противном случае (т. е. если пары вида «глагол — существительное» не допускают никаких диагностических трансформаций) решение не может быть получено.

В соответствии с этим в парах *to lack* — «нуждаться» → *lack* — «нужда»; *to promise* — «обещать» → *promise* — «обещание»; *to support* — «поддерживать» — *support* — «поддержка»; *to love* — «любить» → *love* — «любовь» и т. п. глагол является исходным, а существительное — производным словом; основание — диагностическая трансформация 4, ср. *He lacks money* — «Он нуждается в деньгах» → *his lack of money* — «его нужда в деньгах»; *He promised support* — «Он обещал поддержку» → *his promise of support* — «его обещание о поддержке»; *He supported the president* — «Он поддерживал президента» → *his support of the president* — «его поддержка президента»; *He loved people* — «Он любил людей» → *his love for people* — «его любовь к людям» и т. п. То же самое направление производности обнаруживаем в парах *to live* — «жить» → *life* — «жизнь»; *to dream* — «видеть во сне» → *dream* — «сон»; *to progress* — «прогрессировать» → *progress* — «прогресс»; *to run* — «бежать» → *run* — «бег» (трансформация 5) и т. д.

Тем же способом можно в принципе решить вопрос о направлении производности для пар слов (разумеется, не только английского языка), связанных чередованием звуков, чередованием ударения и другими безаффиксальными

словообразовательными отношениями. Попытаемся сформулировать два ключевых шага этого способа в общем виде: 1) в парах слов, отличающихся аффиксом, слово без этого аффикса является исходным, а слово с аффиксом — производным; 2) если в паре слов, отличающихся аффиксом, C_1 — исходное, а C_2 — производное, и они связаны некоторой диагностической трансформацией T , то в паре слов C_1 и C_2 , не отличающихся друг от друга аффиксами, но связанных той же диагностической трансформацией, слово C_1 является исходным, а слово C_2 — производным.

Перейдем к синтаксису. Мы расскажем о двух экспериментальных методах получения синтаксической информации, которые были первоначально разработаны в рамках исследовательских моделей, хотя с наибольшим успехом они были применены позднее в различных моделях речевой деятельности (см. часть IV). Первый из них — метод непосредственно составляющих (НС) — был разработан Л. Блумфильдом <253>, Р. Уэллсом <375>, З. Харрисом <289>, <295> и другими исследователями¹; второй — трансформационный метод — был разработан в особенности З. Харрисом <195>, <292> и Н. Хомским <199>, <200>². Трансформационный метод (ТМ) возник как дополнение к методу НС, а метод НС был обобщением и уточнением традиционных методов анализа синтаксической структуры предложения; поэтому краткий разбор последних поможет нам понять существо новых методов.

Основные понятия, с помощью которых обычно описывается структура простого предложения, — это понятия

¹ См., например, <44>, <309>. Разновидности этого метода были формализованы И. Бар-Хиллелем <243>, Н. Хомским <199>, В. Ингве <381>, Т. Н. Молошной <122>. В европейской лингвистической традиции был развит весьма близкий к нему метод описания синтаксиса в терминах зависимостей (А. М. Пешковский <143>, <144>, А. де Гроот <284>, Ш. Балли <17>, Л. Теньер <365>, Е. Курилович <91>). Идея «грамматики зависимостей» были формализованы И. А. Мельчуком <116>, С. Я. Фитиаловым <189>, Д. Хейсом <298> и другими исследователями.

² Среди наиболее значительных их предшественников мы бы хотели назвать О. Есперсена (учение о рангах) <55>, А. Фрея (учение о синтаксической транспозиции и конверсии) <273>, Ш. Балли (учение о синтаксической транспозиции) <17>, Е. Куриловича (учение о синтаксической деривации) <89>, В. Н. Сидорова и И. С. Ильинскую <161> (см. стр. 153, где последняя работа кратко изложена).

члена предложения и типа синтаксических отношений.

1. Средством для разграничения членов предложения являются типы вопросов, которые можно задать о предмете или явлении, обозначенном данным словом. Этого средства в общем случае недостаточно для однозначного анализа. Так, в предложении *Жизнь в деревне однообразна* словоформа *в деревне* может рассматриваться и как обстоятельство (*ж и з н ь г д е?*), и как определение (*к а к а я ж и з н ь?*), и как дополнение (*ж и з н ь в ч е м?*). Трудности возникают и при разграничении второстепенных и главных членов предложения <168>, <226>.

Вершиной предложения («абсолютным определяемым») признается обычно подлежащее, а сказуемое рассматривается как зависимый от него элемент, хотя во многих языках сказуемое обладает большей самостоятельностью, чем подлежащее. Это проявляется в том, что сказуемое нельзя опустить без нарушения правильности предложения, в то время как многие типы предложений без подлежащего являются нормой, ср. *Приходите к нам; В народе говорят разное; Мне холодно; Дрожишь?* и т. п.¹.

Наконец, анализ по членам предложения не создает естественной основы для представления предложения в виде иерархии элементов (например, дерева или скобочной записи). Правильным анализом группы дополнения в английском предложении *I saw a first rate second hand book shop* <320, 93> — «Я видел первоклассный букинистический магазин» (буквально «Я видел первого сорта второй руки книжный магазин») является следующий анализ: $\{(first\ rate)\ [(second\ hand)\ (book\ shop)]\}$. Между тем в самих понятиях дополнения и определения (а данная группа других членов предложения не содержит) нет ничего, что бы вынуждало нас анализировать предложение именно таким образом.

2. Выделяются следующие типы синтаксических отношений: предикативные и непредикатив-

¹ Ср. характеристику глагола как «организующего» <30, 651> или «конструктивного» <91> центра предложения, к которому сходятся имеющиеся в последнем синтаксические связи; по словам Ш. Балли, «в индоевропейских языках всякое грамматическое отношение является глагольным. Вся грамматика в целом заключена в глаголе» <17, 120>.

ные, сочинительные и подчинительные; подчинительные отношения делятся, не всегда однозначным образом, на согласование, управление и примыкание.

С логической точки зрения сочинительные и подчинительные отношения совершенно несравнимы. Сочинение — это логическая связка, а не отношение, а подчинение действительно является бинарным отношением в самом строгом смысле слова.

Между типами подчинительных отношений и видами членов предложения существует некоторая зависимость, но нет однозначной связи. Однако характер этой зависимости детально не изучен, и поэтому неясно, каким образом следует переходить от одного анализа к другому.

В отличие от этого метод НС основан на следующих простых содержательных допущениях: 1) существенную роль в синтаксической структуре предложения играет одно-единственное отношение — отношение подчинения; 2) предложение не собирается непосредственно из словоформ, но строится последовательно иерархически: элементарные его части соединяются в простые «строительные блоки», из этих блоков составляются более крупные блоки, пока все предложение не будет представлено в виде единого блока; поскольку отношение подчинения бинарно, каждый блок состоит не более чем из двух частей¹; 3) вершиной иерархии является сказуемое (если считается, что оно подчиняет подлежащее, но не подчинено ему) или группа подлежащего и сказуемого (если считается, что сказуемое подчиняет подлежащее и одновременно подчинено ему)². Легко заметить, что при

¹ С этой точки зрения однородные члены следует рассматривать в качестве единой части блока; ср. <17, 116>.

² Ср. мнение Ш. Балли <17>, Л. Ельмслева <53>, И. И. Ревзина <152> о том, что подлежащее и сказуемое связаны отношением взаимозависимости. С наибольшей глубиной эту точку зрения выразил Е. Курилович <91>, усматривавший в предложении два вида подчинительных отношений: атрибутивные и конститутивные. В словосочетании атрибутивные и конститутивные отношения однонаправлены (конституирующий элемент словосочетания является одновременно определяемым, ср. *красный перец*), а в предложении они разнонаправлены (конституирующий элемент — сказуемое — является не определяемым, а определяющим, ср. *перец красен*). В работе С. К. Шаумяна сходные идеи послужили основой оригинального синтаксического исчисления <208>, <211>. Ср. идеи Л. Блумфильда об эндоцентрических и экзоцентрических конструкциях (часть I, глава 3).

таким подходе отпадает надобность в двух различных анализах. В ходе одного анализа устанавливаются и единицы предложения, и отношения между ними.

Рассмотрим подробнее основные понятия этого метода. Возьмем предложения данного текста и сопоставим каждому из них конструкцию, то есть цепочку символов классов словоформ, например:

Мой друг читает очень интересную книгу

$A_n \quad N_n \quad V \quad D \quad A_a \quad N_a$

Вслед за О. С. Кулагиной <88> будем считать конструкцию правильной, если существует хотя бы одно правильное предложение, которому она может быть поставлена в соответствие. В частности, приведенная выше конструкция правильна хотя бы потому, что она соответствует правильному предложению *Мой друг читает очень интересную книгу*. Для дальнейших определений воспользуемся понятием ранга, близким к тому, которое предложено А. В. Гладким <43а>. Будем рассматривать произвольную пару символов XU как синтагму первого ранга, если в любой содержащей ее правильной конструкции она может быть заменена, без нарушения правильности конструкции, символом Y , а символ Y может быть заменен, на тех же условиях, парой XU . При некоторых условиях синтагмой первого ранга можно считать, например, пару символов DA_n (ср. *очень хороший человек*), заменимую на A_n (ср. *хороший человек*), и наоборот. Правда, от наречий могут зависеть другие наречия (ср. *довольно хорошо скроенный костюм*), а относительные прилагательные наречий не подчиняют, но эти случаи тоже допускают формальное описание и здесь для простоты не учитываются. Назовем синтагмой i -го ранга пару символов XU , если в любой правильной конструкции, где X не входит в состав синтагмы $(i-1)$ -го ранга, она заменима символом Y , и наоборот. Будем говорить, что X и Y — непосредственно составляющие синтагмы, причем Y — ядро, а X — маргинальный (зависимый) элемент.

Используя понятия синтагмы и непосредственно составляющей, мы можем установить синтаксические связи между словоформами и иерархию связей. Действительно, синтагмами в указанном выше смысле являются пары

$A_n N_n$, DA_a , $N_n V$ и пр., причем во всех этих парах главным является второй элемент, а зависимым — первый. Поскольку, как следует из данных выше определений, каждая синтагма по своим внешним синтаксическим свойствам эквивалентна ядру, взятому в отдельности, мы можем записать следующие правила «свертывания» синтагм по НС (S — символ предложения):

- | | |
|---|---|
| <p>(1) $D + A_a \rightarrow A_a$
 $D + A_n \rightarrow A_n$
 и вообще
 $D + A_x \rightarrow A_x$</p> | <p>(2) $A_n + N_n \rightarrow N_n$
 $A_a + N_a \rightarrow N_a$
 и вообще
 $A_x + N_x \rightarrow N_x$</p> |
| <p>(3) $V + N_a \rightarrow V$
 $V + N_n \rightarrow V$
 и вообще
 $V + N_x \rightarrow V$</p> | <p>(4) $N_n + V \rightarrow S$
</p> |

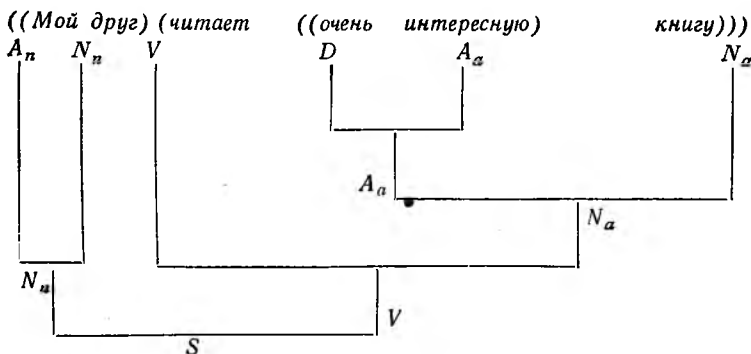
Нетрудно заметить, что между этими правилами имеется определенная связь; так, правило (1) $D + A_x \rightarrow A_x$ связано с правилом (2) $A_x + N_x \rightarrow N_x$, а правило (2) $A_x + N_x \rightarrow N_x$ — с правилом (3) $V + N_x \rightarrow V$: в обоих случаях результирующий элемент предыдущей синтагмы является НС следующей синтагмы; это значит, что более простая синтагма может вкладываться в более сложную, выступая в качестве НС последней. Порядок вложения можно изобразить скобками; запись $(D + A_x) + N_x$ показывает, что правило (1) предшествует правилу (2). Очевидно, что этот порядок однозначно вытекает из описанной выше процедуры: существуют правильные цепочки, содержащие синтагму $A_x + N_x$ (ср. *интересная книга*), в которые может быть вложена синтагма $D + A_x$ (ср. *очень интересная книга*), но не существует правильных цепочек вида $*D + N_x$ (ср. **очень книга*), которые мы должны были бы предположить, если бы считали, что правило (2) предшествует правилу (1). Таким образом на множестве правил НС устанавливается определяемое рангом синтагм отношение порядка.

Выработанная таким образом упорядоченная последовательность правил является выходом собственно экспериментальной части модели НС. Затем (ср. алгоритм Б. В. Сухотина) ее можно применить для анализа предложений некоторого текста по НС. Анализ по НС называется *свертыванием*. Из сделанных выше замечаний следует, что он подчиняется следующим ограничениям:

1) одновременно можно свертывать не больше двух символов — НС некоторой синтагмы — в один новый; 2) запрещается перестановка символов; 3) порядок применения правил фиксирован. Весь процесс анализа предложения (свертывания по НС) может быть изображен тремя способами: в виде правил подстановки (самый информативный способ), в виде скобочной записи и в виде дерева НС. Так, структура предложения *Мой друг читает очень интересную книгу* адекватно описывается следующими пятью правилами:

Правило	Результат применения
(1) $D + A_x \rightarrow A_x$	$A_n N_n V D A_a N_a$
(2) $A_x + \bar{N}_x \rightarrow \bar{N}_x$	$A_n N_n V A_a \bar{N}_a$
(3) $A_x + N_x \rightarrow N_x$	$A_n N_n V N_a$
(4) $V + N_x \rightarrow N_x$	$N_n V N_a$
(5) $N_n + \bar{V} \rightarrow S$	$N_n V$ S

Скобочная запись эквивалентна записи структуры предложения в виде дерева НС:



Помимо весьма схематически изложенной здесь процедуры анализа предложения по НС, существует несколько иная процедура, впервые намеченная Л. Блумфильдом <253> и получившая законченную формулировку в классической работе Р. Уэллса <375>. В соответствии с правилами Л. Блумфильда и Р. Уэллса предложение, записанное в виде последовательности морфем, сначала разбивается границей на две наиболее крупные части; затем каждая из полученных частей делится на две новые и т. д., пока мы не дойдем до отдельных морфем, которые считаются

конечными составляющими. С. Чэтмен <260> разъяснил впоследствии, что на каждом шаге анализа границы проводятся в местах максимальной независимости непосредственно составляющих друг от друга. Независимость измеряется вероятностью появления некоторой части после данной части. Независимость тем больше, чем меньше эта вероятность. По этому правилу из двух разбиений предложения *он-а | чит-а-л-а* и *он-а чит-а-л | -а* мы должны выбрать первое, так как вероятность появления части *чит-а-л-а* после части *он-а* весьма невелика (части независимы), а вероятность появления части *-а* после части *он-а чит-а-л* равна единице (последняя часть целиком зависит от первой).

Выше мы сказали, что метод НС был дополнен трансформационным методом. Посмотрим, чем была вызвана необходимость в такого рода дополнении.

В разное время в специальной литературе <181>, <199>, <309>, <359>, <163> отмечались следующие недостатки метода НС:

1) В ряде случаев метод НС приводит к слишком грубым решениям, поскольку он не дает возможности фиксировать различия в синтаксической структуре предложений, интуитивно достаточно очевидные. Так, один и тот же анализ по НС приписывается следующим двум предложениям:

Изучение событий становится интересным

Развитие событий становится интересным



Между тем эти предложения построены не вполне одинаково: в первом случае события являются объектом некоторого действия (кто-то изучает события); во втором случае события являются субъектом действия (они сами развиваются). Анализ по НС, как и традиционный анализ по членам предложения, не вскрывает этих структурных различий.

2) Одним из критериев адекватности синтаксического метода является его способность предоставить на выбор несколько различных путей анализа для синтаксически и семантически двусмысленных предложений. Рассмотрим с этой точки зрения предложение *Сплочение рабочих бригад вызвало осуждение товарища министра*. Оно допускает

два различных анализа по НС: ((*Сплочение (рабочих бригад)*)) (*вызвало (осуждение (товарища министра))*))) и (((*Сплочение (рабочих бригад)*)) *вызвало*) (*осуждение (товарища министра)*)). Между тем в действительности это предложение допускает 32 различных синтаксических анализа (и 32 различных осмысления), и, следовательно, анализ по НС вскрывает лишь одну шестнадцатую часть реальных возможностей (настойчивый читатель сумеет сам увидеть все эти возможности).

3) Анализ по НС не показывает связи между активными и пассивными, утвердительными и отрицательными, утвердительными и вопросительными конструкциями, которые по ряду веских причин хотелось бы рассматривать как связанные отношением производности.

4) С трудом поддаются анализу по НС некоторые типы экзоцентрических конструкций (см. стр. 43), конструкции с так называемыми разрывными морфемами и, наконец, конструкции с аппозитивными элементами вида *товарищ Иванов, Иоанн Креститель, озеро Байкал, пароход «Революция», князь Андрей* (примеры из <144, 295—296>).

5) В случае, если анализируемое предложение достаточно сложно, деревья, скобочная запись и правила подстановки приобретают труднообозримый вид. По-настоящему эффективным метод НС является в рамках простого предложения.

Отметим, наконец, что метод НС не дает возможности ответить на вопрос о том, что является элементарной единицей языка на синтаксическом уровне анализа, и проследить историю образования сложного синтаксического типа из простых. Именно это обстоятельство и явилось непосредственной причиной разработки трансформационного метода.

ТМ возник первоначально как продолжение дистрибутивных процедур на синтаксическом уровне.

Как мы помним, одна из основных задач всякой модели исследования состоит в том, чтобы найти элементарную единицу языка на каждом из выделяемых уровней. На фонологическом уровне — это фонема, на морфологическом — морфема. З. Харрис предположил, что основной единицей синтаксического уровня является некоторый тип простого предложения (конструкция, поставленная в соответствие предложению). Мы помним, далее, что в моделях исследования каждая единица языка, получаемая на

выходе, определяется как класс дистрибутивно эквивалентных единиц текста, которые считаются ее вариантами. Фонема — это класс аллофонов, которые находятся в дополнительной дистрибуции или свободном чередовании; морфема — это класс так же распределенных алломорфов (см. стр. 50). По-видимому, и тип предложения является классом вариантов, реализуемых в тексте; однако отличие типа предложения от всех прочих языковых единиц состоит в том, что он не допускает дистрибутивного определения. Понятие дистрибуции предложения, или дистрибуции типа предложения, практически бессодержательно, так как на окружение предложения (типа предложения) не накладывается никаких синтаксических ограничений: справа и слева от данного предложения могут стоять предложения любой синтаксической структуры. Это значит, что все предложения (и типы предложений) имеют одну и ту же дистрибуцию и, следовательно, по дистрибутивным признакам не могут быть отличены друг от друга. В этих условиях З. Харрис <195>, <292> и предпринял попытку использовать для отождествления предложений (и типов предложений) их трансформационные признаки.

Мы начнем обсуждение ТМ с изложения двух его основных идей. Первая идея состоит в том, что синтаксическая система языка может быть разбита на ряд подсистем, из которых одна является ядерной, исходной, а все другие — ее производными. Ядерная подсистема — это набор элементарных типов предложений; любой сколько-нибудь сложный синтаксический тип представляет собой трансформ одного или нескольких ядерных типов, т. е. известную комбинацию ядерных типов, подвергнутую ряду преобразований (трансформаций). Это представление можно пояснить сравнением синтаксической системы языка с таблицей Менделеева: громадное число «молекул» (сложных синтаксических типов) описывается с помощью небольшого числа «элементов» (ядерных типов) и небольшого числа «реакций» (трансформаций).

Изложенный здесь взгляд на синтаксическую систему языка как иерархию, в основе которой лежит ограниченное множество простейших синтаксических типов, подтверждается некоторыми весьма интересными типологическими и историческими данными, данными детского языка и языка афатиков. Оказывается, что сложные синтаксические типы являются продуктом длительного разви-

тия, в частности развития письменности; они позднее всего усваиваются ребенком и легче всего разрушаются при афазиях. В противоположность этому простейшие (ядерные) синтаксические типы, наиболее непосредственным образом описывающие внеязыковую ситуацию, обнаруживают, как и положено фундаменту системы, сохраняемость на протяжении весьма значительных исторических периодов, устойчивость в синхронном состоянии и универсальность. Рассмотрим некоторые примеры.

Было замечено, что «языки гораздо больше похожи друг на друга в отношении простейших конструкций..., чем в отношении структуры взятых целиком предложений» (245, 7). Чем конструкция сложнее, тем больше вероятность того, что она окажется специфической для данного языка. Таков *ablative absolutus* в латыни, *genitive absolute* в древнегреческом, *dative absolute* в древнерусском языке и т. д. Специфическими для германских и романских языков являются сложные конструкции с инфинитивом и причастием типа *nominative cum infinitivo*, *accusative cum infinitivo* и другие подобные. Даже такая, казалось бы, несложная конструкция, как *NcopAN* (ср. *Сократ был мудрый афинянин*) или *NVAN* (ср. *Он увидел высокого мужчину*), отсутствует в ряде языков, а соответствующая мысль выражается в них двумя более простыми предложениями. Таковы, по данным С. Ньюмена (327) и Дж. Гринберга (373), йокуцкий и арапешский языки, в которых высказывание, эквивалентное предложению *Он увидел высокого мужчину*, разбивается на два отдельных высказывания: *он увидел мужчину* и *мужчина был высок*. Таким образом, можно подозревать прямую зависимость между универсальностью (неспецифичностью) конструкции и ее «ядерностью»; фундамент синтаксической системы, т. е. множество простейших типов, практически не меняется от языка к языку.

Различные типы сложноподчиненных предложений исторически развиваются из сочинения, а сложносочиненные предложения в свою очередь возникают из простого соположения двух или более предложений ядерного вида¹ (320, 351). Отметим, например, что сложноподчиненное предложение с придаточным определительным — позд-

¹ В данном случае синхронный трансформационный анализ (см. ниже) соответствует преобразованиям, имевшим место в диахронии; ср. (60), (286).

нейшее явление; более древним способом связи является бессоюзное соединение двух простых предложений, ср. *И отдали меня Афонасью Пашкову в полк — людей с ним было 600 человек* <26, 463> (сейчас мы бы сказали... в котором с ним было 600 человек людей). В говорах этот более древний способ связи предложений сохранился до сих пор, ср. *Где дели тряпку, полы мыли?* (т. е. *которой полы мыли?*) <26>.

Весьма интересны данные детского языка. Первым типом неаморфных предложений, который усваивается ребенком в возрасте около 1 года 10 месяцев, являются утвердительные двусоставные двусловные предложения типа *Собака бежит*. Повелительные, восклицательные, вопросительные и отрицательные предложения (трансформы ядерных) появляются позже утвердительных; безличные — позже личных. Предложения с дополнениями (*ты купила книжку*) появляются на три месяца раньше предложений с согласуемыми определениями. Сложное предложение и предложение с однородными членами усваиваются позже простого (в возрасте около 2-х лет), причем до 2 лет 3 месяцев оно остается бессоюзным. Простейшая форма, в которой усваивается сложное (сочиненное) предложение, — форма перечисления, ср. *Мама плетет, Баба плетет; Собака сидит, Девочка сидит* <43>.

Если у ребенка усвоение синтаксической системы языка начинается с ядерных типов, от которых он продвигается ко все более сложным трансформам, то у афатика разрушение синтаксической системы начинается с трансформов и заканчивается ядерными типами <71>. Восстановление синтаксической системы идет с большим трудом и начинается, как и следует ожидать, с ядерных типов. Усвоив двусловную предикативную синтагму типа *Корабли стоят*, афатик долгое время не может овладеть трансформацией, присоединяющей к этой синтагме адвербиальное пояснительное слово или предложную конструкцию <71, 85>.

Все эти факты свидетельствуют о реальности ядерной системы, которая постоянно дает о себе знать вопреки тому обстоятельству, что предложения ядерных типов крайне редко встречаются в непринужденном диалоге и почти не встречаются в письменных текстах.

Вторая важная идея трансформационного метода состоит в том, что ядерные предложения (и их трансформы)

связаны с элементарными ситуациями и, следовательно, ядерные типы (вместе со своими трансформами) — с классами элементарных ситуаций. Так, тип, который задается ядерными предложениями вида $N_n^1 V N_a^2$ (в русском языке) и трансформом $N_n^2 V$ ся от N_g^1 , связан с ситуацией, которую можно содержательно описать как ситуацию «нарушения состояния физического равновесия», ср. *Взрыв гнет (ломает) рельсы* → *Рельсы гнутся (ломаются) от взрыва*; *Ветер шатает (сотрясает) дом* → *Дом шатается (сотрясается) от ветра*. Тип, который задается ядерными предложениями вида $N_n^1 V N_a^2$ и трансформом $N_n^2 V$ ся к N_a^1 , связан с ситуацией «каузации движения объекта по направлению к субъекту», ср. *Магнит притягивает железо* → *Железо притягивается к магниту*; *Свет тянет мотылька* → *Мотылек тянется к свету*; *Капитан возвращает роту* → *Рота возвращается к капитану*. Тип $N_n^1 V N_a^2 \leftrightarrow N_a^2 V$ ся за N_a^1 связан с ситуацией «нарушения состояния эмоционального равновесия», ср. *Сын беспокоит (волнует, огорчает, радует, тревожит) мать* ↔ *Мать беспокоится (волнуется, огорчается, радуется, тревожится) за сына*. Тип $N_n^1 V N_a^2 \leftrightarrow N_n^2 V$ ся на N_a^1 связан с ситуацией «каузации отрицательных эмоций», ср. *Он злит (обижает, сердит) меня* ↔ *Я злюсь (обижаюсь, сержусь) на него* <9>. Говоря словами Н. Хомского, трансформационный анализ позволяет, в частности, «свести вопрос о том, как понимается язык, к вопросу о том, как понимаются ядерные предложения... Поскольку ядерные предложения, лежащие в основе данного высказывания, передают, по-видимому, главное в содержании высказывания, такой анализ как будто предоставляет средства для исследования строения связного текста; до сих пор эта задача оставалась за пределами лингвистического анализа» <263, 291>.

Перейдем к более систематическому рассмотрению ТМ, который в моделях исследования часто именуется трансформационным анализом.

Трансформационный анализ получает на входе текст и множество правильных фраз с уже составленными дугами непосредственных синтаксических связей и вырабатывает на выходе информацию о ядерных типах (или предложениях) данного языка и трансформа-

циях, действующих в его синтаксической системе. Чтобы получить эту информацию, необходимо каждое входящее в «машину» предложение свести к одному или нескольким ядерным, указав при этом, какие трансформации и в каком порядке воздействовали на него.

Синтаксическое понятие ядерного типа или ядерного предложения сравнимо с лексико-графическим понятием основной формы слова. Подобно тому как в множестве форм существительного форма именительного падежа единственного числа выбирается (например, в словаре) в качестве основной, хотя и не единственной, представительницы данного существительного, ядерное предложение — основной, но не единственный вариант, в котором манифестируется некоторое простое предложение; оно входит в класс трансформов простого предложения наряду с другими его трансформами.

Простое предложение, как таковое, существует в двух видах и, следовательно, может быть задано двумя способами. Во-первых, оно может быть задано своим «корнем» или «основой»¹, т. е. основами всех входящих в него словоформ с указанием синтаксических отношений между ними, ср.


профессор — осмотр — пациент —,

(например, *профессор осматривает пациента*). Если у предложения, как и у слова, имеется корень (или основа), то, очевидно, у него должна быть и парадигма. Этот вывод, имплицитно содержащийся в работах З. Харриса <195> и Н. Хомского <199>, был вполне четко сформулирован Д. Уорсом <379>, который первым начал разработку синтаксических парадигм (парадигм предложения)². В парадигму предложения входят, в частности, его образовательные и изменительные формы (аналогия с формами словообразования и словоизме-

¹ В литературе для обозначения этого понятия используется несколько громоздкий термин «лежащая в основе цепочка» (the underlying string Н. Хомского <199>). Простой и наглядный термин «корень» использовался (устно) К. И. Бабицким.

² Независимо от трансформационной грамматики оригинальная и богатая мыслями концепция парадигм предложения была развита Н. Ю. Шведовой <217>.

нения); к числу «образовательных» форм предложения относятся, например, так называемые номинализации, ср. *осмотр пациента профессором; то, что профессор осматривает пациента; профессор, осматривающий пациента* и т. д.; к числу «изменительных» форм — формы времени и наклонения, вопроса и отрицания и т. д.¹.

Следовательно, второй способ адекватно задать предложение состоит в фиксации его парадигмы или указании множества его форм (вариантов или трансформов), ср.

- (1) *Профессор осматривает пациента.*
- (2) *Осматривает ли профессор пациента?*
- (3) *Профессор не осматривает пациента.*²
- (4) *Пациент осматривается профессором.*
- (5) *Профессор будет осматривать пациента.*
- (6) *Пациент будет осмотрен профессором.*
- (7) *Кем будет осмотрен пациент?*
- (8) *Кем был осмотрен пациент?*
- (9) *Кем осмотрен пациент?*
- (10) *Осмотр пациента профессором.*
- (11) *Профессор, осматривающий пациента.*
- (12) *Пациент, осмотренный профессором, и т. д.*

Ядерным называется простейший из всех трансформов данного предложения, т. е. трансформ, который 1) имеет форму предложения и 2) содержит наименьшее число грамматических морфем. В терминах «категорий» этот трансформ есть предложение в изъявительном наклонении, настоящем времени, активном залоге, утвердительной форме, не содержащее модальных и эмфатических слов и т. д. <199>. Он образуется от корня предложения с помощью так называемых **о б я з а т е л ь н ы х** трансформаций и считается исходным элементом системы; все остальные трансформы данного предложения рассматриваются в качестве производных элементов, образованных от исходного (или от корня) с помощью факультативных трансформаций. В нашем примере ядерным является первый трансформ (*Профессор осматривает пациента*); трансформ (2) образован от него с помощью вопроситель-

¹ См. <217>, <362>, где морфемы времени и наклонения трактуются как морфемы предложения.

² Утвердительные и отрицательные предложения «имеют взаимно противоположный смысл, но им соответствует одна и та же действительность» <38>.

ной трансформации; трансформ (3) — с помощью отрицательной; трансформ (4) — с помощью пассивизации и т. д.

Эти трансформации являются элементарными синтаксическими преобразованиями. Однако трансформы могут быть результатом не только элементарного, но и сложного, т. е. двойного, тройного и вообще n -кратного, преобразования ядерного предложения (или его корня); так, трансформ (6) образован в результате воздействия на ядерное предложение трансформации пассивизации и трансформации будущего времени, а в трансформе (7) обнаруживается действие пассивизации, трансформации будущего времени и вопросительной.

Как легко заметить, здесь наше понимание трансформации несколько изменяется. Если раньше трансформацией мы называли пару фраз, отвечающих определенным условиям, то теперь мы рассматриваем трансформацию как некоторую операцию, переводящую одну фразу в другую по определенным правилам. Будучи операцией, каждая трансформация может быть описана вполне строго в терминах тех элементарных изменений в исходной фразе, которые необходимо произвести, чтобы из исходной фразы получить данный трансформ <180>. Так, трансформация будущего времени для глаголов несовершенного вида может быть описана следующим образом: введи в предложение глагол *быть* в будущем времени в том числе и лице, в котором стоит личный глагол исходного предложения; замени личный глагол инфинитивом.

Для каждой трансформации можно установить ей обратную (ср. <14>, <195>, <199>). Так, для описанной выше трансформации будущего времени T_6 можно установить обратную трансформацию T_6^{-1} , переводящую фразу *Профессор будет осматривать пациента* во фразу *Профессор осматривает пациента* по следующим правилам: замени инфинитив личным глаголом настоящего времени в том числе и лице, в котором стоит вспомогательный глагол *быть*; опусти вспомогательный глагол *быть*. Любое предложение, каким бы сложным оно ни было, может быть сведено к ядерному предложению (или предложениям) с помощью обратных трансформаций.

Мы обнаруживаем, таким образом, что если установлено множество трансформов данного предложения, то мы можем решить обе поставленные выше задачи трансформационного анализа: получить информацию о ядерных

предложениях (типах) и трансформациях, действующих в синтаксической системе данного языка. Следовательно, вопрос сводится к тому, как установить множество трансформов данного предложения.

На первый взгляд кажется, что сделать это достаточно просто; мы располагаем определением отношения трансформируемости (см. стр. 155), которое является эквивалентностью и в качестве такового порождает на множестве фраз данного языка разбиение (деление на непересекающиеся классы); каждый класс разбиения и является искомым множеством трансформов.

В действительности, однако, ситуация является более сложной. Рассмотрим, например, трансформы (6) и (7). Они имеют одно и то же дерево синтаксических связей и одни и те же лексические морфемы, за тем важным исключением, что в (7) вместо существительного *профессором* стоит вопросительное слово *кем*. Здесь как будто нарушается первое условие трансформируемости — требование тождества лексических морфем. Аналогичная ситуация имеет место во многих других случаях, например в случае определительных придаточных предложений, ср. *Он встретил друга, которого он не видел 20 лет*. По-видимому, это сложное предложение образовано из двух простых: (1) *он встретил друга*; (2) *он не видел друга 20 лет*. Однако придаточное определительное предложение и предложение (2), из которого оно, по нашему предположению, образовано, не отвечают первому условию трансформируемости, т. е. условию идентичности лексических морфем, поскольку в трансформе *которого он не видел 20 лет* опущено слово *друга*.

В рассмотренных выше случаях в трансформе оставался хотя бы след опущенной лексической морфемы (ср. *кем, которого*). Еще более трудными для трансформационного анализа являются фразы, в которых от опущенного слова не остается даже следа.

Типовыми случаями являются:

1) Безличные предложения типа *Про книгу было забыто <159>*, которые кажется естественным связать с личными предложениями типа *Некто забыл про книгу*.

2) Сложносочиненные предложения двух типов: а) *Некоторые товарищи пришли вовремя; другие опоздали* (из *Другие товарищи опоздали*); б) *Он говорил там, а не я* (из *Я не говорил*).

3) Сравнительные предложения типа *Он выше, чем я* (из *Я высок*).

4) Предложения, которые можно было бы назвать скрытыми аналогами сложных предложений. К их числу относятся: а) предложения с прилагательным в атрибутивной функции, ср. *Сократ был мудрый афинянин* (по предположению, трансформ двух предложений: *Сократ был мудр* и *Сократ был афинянин*); б) предложения с однородными членами, ср. *Ермолай сидел около меня, дремал и клевал носом* (по предположению, трансформ трех предложений: *Ермолай сидел около меня. Ермолай дремал. Ермолай клевал носом*); в) предложения с любыми неличными формами глагола, ср. *Ермолай дремал, сидя около меня, и клевал носом*. Среди них особый интерес представляют предложения с субъектным и объектным инфинитивом, ср. *Я хочу пить* из *Я хочу, чтобы я пил*; *Я обещаю ему прийти* из *Я обещаю, что я приду* и *Я велел ему прийти* из *Я велел ему, чтобы он пришел*; *Я просил его прийти* из *Я просил его, чтобы он пришел*; г) предложения с субъектным и объектным предикативом, ср. *Он вернулся солдатом* из *Он вернулся* и *Он — солдат*; *Мы избрали его президентом* из *Мы избрали его* и *Он — президент* и т. п.

Во всех перечисленных случаях сведение встретившихся в тексте предложений к ядерным сопряжено с необходимостью восстановить в них части, которые, по нашему предположению, были опущены в процессе построения предложения. Этот вопрос З. Харрис предлагает решить расширением обычных представлений о формах существования лексических морфем <195>. С его точки зрения, каждая лексическая морфема есть множество вариантов, в число которых входит п р о м о р ф е м а¹, встречающаяся во фразах типа *Он встретил друга, которого он не видел 20 лет* или *Кем будет осмотрен пациент, и н у л е в о й* в а р и а н т, встречающийся во фразах типа *Он выше, чем я (высок)* или *Он говорил там, а не я (говорил)*.

К числу проморфем относятся: 1) *кто, что, который, когда, где, он, она, это* для слов класса *N*; 2) *этот, некоторый, другой, какой* и др. для слов класса *A*; 3) *do, will* (в английском языке) для слов класса *V* и т. д.

¹ Необходимость в понятии такого типа была отчетливо сформулирована еще в работах Ф. Брюно <257, 173> и А. Фрея <273, 113>; см. также <309, 257—258>.

Восстановление во фразе, хотя бы с точностью до класса, тех лексических морфем, которые представлены в ней в виде проморфем и нулей, делает возможным трансформационный анализ в указанном выше смысле. Каждое сложное, в том числе и неочевидно сложное, предложение представляется в виде набора простых предложений; устанавливаются трансформации, обратные тем, с помощью которых данное сложное предложение могло быть построено из простых. Затем каждое простое предложение преобразуется таким образом, чтобы его трансформ содержал возможно меньшее число грамматических морфем и чтобы при этом были соблюдены оба сформулированные выше условия трансформируемости. В ходе этой процедуры фиксируются все (обратные) трансформации, пока мы не дойдем до предложения, не допускающего дальнейших упрощений и, следовательно, являющегося ядерным. На последнем этапе мы можем получить ядерный тип, заменив каждую словоформу ядерного предложения символом класса словоформ, к которому она принадлежит. Таким образом З. Харрис получил для английского языка ядерные типы *NV*, *NVPN*, *NVN*, *N is N*, *N is A*, *N is PN*, *N is D*, *N is P N* и ряд других <195>.

Попытаемся теперь кратко обозреть то множество трансформаций, которое мы получаем в ходе описанного выше процесса сведения реальных предложений к ядерным типам¹. Перечислить все трансформации даже для одного языка мы здесь не сможем и поэтому ограничимся указанием основных типов трансформаций (нумерация классов ведется в двоичной системе): 0. Трансформации, охватывающие более одного предложения ядерного вида; 1. Трансформации, не выходящие за рамки одного предложения; 00. Явно сложные трансформации; 01. Неявно сложные трансформации; 000. Трансформации с сочинением; 001. Трансформации с подчинением; 0000. Трансформации без нулевых лексических морфем, ср. *он поет, и я пою*; 0001. Трансформации с нулевыми лексическими морфемами, ср. *он поет, а не я*; 0010. Трансформации без нулевых лексических морфем, ср.

¹ Любопытная идея экспериментальной процедуры получения трансформаций высказывается в работе Д. Уорса <180>. В работах С. К. Шаумяна и П. А. Соболевой <208>, <211> построено исчисление трансформаций, позволяющее получить все типы трансформов для любого типа фраз.

они приехали, когда поезд ушел; я не верю тебе, потому что однажды ты меня обманул; я верю тебе, хотя однажды ты меня обманул; где раньше зрели хлеба, теперь ничего не растет; 0011. Трансформации с нулевыми лексическими морфемами и проморфемами, ср. он сказал, что придет; он встретил друга, которого не видел 20 лет; 010. Трансформации с неявным сочинением; 011. Трансформации с неявным подчинением; 0100. Трансформации с однородными членами, ср. швед, русский, колет, рубит, режет; 0101. Трансформации с предикативом, ср. он вернулся солдатом, они избрали его президентом, она помнит мужа молодым; 0110. Трансформации с деепричастием, ср. он пел, сидя на лавке; 10. Номинализирующие трансформации, ср. чтение книги мальчиком; мальчик, читающий (который читает) книгу; книга, читаемая (которая читается) мальчиком; то, что мальчик читает книгу; мальчик за чтением книги; 11. Неноминализирующие трансформации; 110. Трансформации с изменением класса слов; 111. Трансформации без изменения класса слов; 1100. Трансформации с обращением элементов, ср. книга стоит рубль ↔ рубль — стоимость книги, я учу его ↔ он — мой ученик; 1101. Трансформации без обращения элементов, ср. книга интересуется меня ↔ ↔ книга интересна мне, я учу его ↔ я — его учитель; 1110. Трансформации с обращением элементов; 1111. Трансформации без обращения элементов; 11100. Пассивные трансформации; 11101. Непассивные трансформации; 111000. Трансформации с нулевыми лексическими морфемами, ср. дом строится; 111001. Трансформации без нулевых лексических морфем, ср. дом строится рабочими, собака пугается звуков, флаги развеваются на ветру; 111010. Симметричные трансформации, ср. треугольник равен квадрату ↔ квадрат равен треугольнику, он спорит (беседует) со мной ↔ я спорю (беседую) с ним, он сравнивает копию с оригиналом ↔ он сравнивает оригинал с копией; 111011. Несимметричные трансформации, ср. царь пожаловал ему шубу ↔ царь пожаловал его шубой; 11110. Обезличивающие трансформации; 11111. Необезличивающие трансформации; 111100. Трансформации с нулевыми лексическими морфемами, ср. про книгу было забыто; 111101. Трансформации без нулевых лексических морфем, ср. дорогу занесло снегом; 111110. Трансформации с сохранением утвердительной формы (эмфатическая, модальная), ср. пришел-то он пришел; понимать — я понимаю, поме-

шал не помещал и другие подобные, ср. <214>, <220>; 111111. Трансформации с изменением утвердительной формы (вопросительная, отрицательная).

В представленной здесь типологии трансформаций не проводилось различия между продуктивными и непродуктивными (грамматическими и лексическими) трансформациями <195>, <261>, <199>. Вопрос этот изучен мало, хотя представляет, по-видимому, выдающийся интерес, особенно для порождающих грамматик, которые либо не должны содержать непродуктивных трансформаций, либо, если они их содержат, иметь для каждой непродуктивной трансформации список элементов, к которым она применима. В частности, для трансформации типа *царь пожаловал ему шубу* ↔ *царь пожаловал его шубой* должно быть указано, что она применима к фразам с глаголами *дарить*, *жаловать*, *обеспечивать* и *сужать* в центральной позиции; для трансформации типа *камни имеют разный вес* ↔ *камни разного веса* должно быть указано, что она применима к исходным фразам с существительными *ум*, *мужество*, *вес*, *сорт*, *сложение*, *рост* в позиции после глагола *иметь* (ср. *человек большого ума*, *женщина редкого мужества*, *мужчина крепкого сложения*), но не применима к фразам с другими существительными, хотя вплоть до начала XIX века это было, по-видимому, возможно, ср. *офицер сурового лица*, *девка немецкой талии*; *стихи дактилического окончания* и т. п. <215>. Следовательно, если мы хотим иметь трансформационные правила порождения фраз указанного здесь типа, мы должны в ходе трансформационного анализа получить, помимо ядерных типов и трансформаций, детальную классификацию лексики и определить область действия каждой трансформации в терминах лексических классов.

Последняя задача трансформационного анализа состоит в том, чтобы у п о р я д о ч и т ь найденные в ходе анализа трансформации. Рассмотренный выше материал показывает, что на множестве трансформаций имеет место отношение частичного (а не строгого) порядка. В предложении *Пациент будет осмотрен профессором* мы можем сначала провести депассивизацию, а затем — трансформацию, обратную трансформации будущего времени, но можем провести эти трансформации и в обратном порядке. Существо дела от этого не изменится, так как на любом шаге анализа мы в любом из двух вариантов будем полу-

чать правильные предложения. Существуют, однако, такие части трансформационной системы, в которых последовательность правил жестко фиксирована. Очевидно, например, что при трансформационном анализе факультативные трансформации, сводящие некоторое предложение к его ядру или ядрам, предшествуют обязательным трансформациям, которые сводят ядро к корню предложения. Фиксированным является порядок и для некоторых других трансформаций. Анализируя предложение *Ответа из полка не пришло*, мы должны сначала провести трансформацию, обратную обезличиванию (с результатом *ответ из полка не пришел*), а затем — трансформацию, противоположную отрицательной (с результатом *ответ из полка пришел*). Аналогичным образом для анализа предложения *Им не требуется документов* применяется последовательность обратных трансформаций *им не требуется документов* ↔ *им не требуются документы* ↔ *им требуются документы* ↔ *они требуют документы* (примеры заимствованы из <159>). Если указанный здесь порядок будет нарушен, то на каком-то шаге анализа мы получим неправильное предложение. Так, если мы поменяем местами трансформации, обратные трансформациям отрицания и обезличивания, мы на первом же шаге анализа получим неправильные предложения * *Ответа из полка пришло*, * *Им требуется документов*.

Установлением отношения порядка на множестве трансформаций задачи трансформационного анализа исчерпываются. В результате мы получаем метод, более сильный, чем метод непосредственно составляющих. Он, в частности, приписывает различные анализы предложениям *Изучение событий становится интересным* и *Развитие событий становится интересным* (ср. стр. 179), показывает связь между активными и пассивными, утвердительными и вопросительными, утвердительными и отрицательными конструкциями. Он, наконец, приписывает несколько различных анализов омонимичным предложениям. В частности, он дает 32 возможных синтаксических решения для предложения *Сплочение рабочих бригад вызвало осуждение товарища министра*, которое по НС анализируется всего двумя способами (см. стр. 180). Секрет этого предложения заключается в том, что оно построено из пяти различных ядерных предложений, причем каждое из них использовано в такой форме, в которой оно омони-

мично другому ядерному предложению: 1) *некто сплачивает рабочих и рабочие сплачиваются* (омонимия в форме номинализации *сплочение рабочих*); 2) *некто сплачивает бригады и бригады сплачиваются* (омонимия в форме номинализации *сплочение бригад*); следовательно, фраза *сплочение рабочих бригад* допускает четыре различных осмысления; 3) *некто осуждает товарища и товарищ осуждает кого-то* (омонимия в форме номинализации *осуждение товарища*); 4) *некто осуждает министра и министр осуждает кого-то* (омонимия в форме номинализации *осуждение министра*); следовательно, фраза *осуждение товарища министра* также допускает четыре различных анализа, причем каждый из них может сочетаться с любым из четырех возможных анализов фразы *сплочение рабочих бригад*; всего, таким образом, получается шестнадцать различных синтаксических решений (и шестнадцать различных осмыслений). Наконец, 5) ядерное предложение N_n^1 вызвало N_a^2 , где оба существительных среднего рода, омонимично предложению N_a^1 вызвало N_n^2 (ср. *весло задело платье; сплочение вызвало осуждение*), что дает два новых решения, каждое из которых способно сочетаться с любым из шестнадцати упомянутых выше анализов. В результате мы получаем тридцать два искомых решения.

Покажем в заключение, каким образом изложенные выше принципы трансформационного анализа применяются для решения тех или иных конкретных вопросов грамматики. В качестве иллюстрации мы рассмотрим уже упоминавшуюся работу Р. Ружички <159>, посвященную трансформационному анализу безличных предложений в русском языке.

Было замечено, что безличные предложения русского языка не являются однородным классом, так как некоторые из них весьма тесно связаны с определенными типами личных предложений. Так, В. А. Богородицкий выводил безличные предложения типа *Морозит* из личных предложений типа *Мороз морозит* (см. стр. 83); В. В. Виноградов связывал безличные предложения типа *Градом побило рожь* с личными предложениями типа *Град побил рожь* <35>; В. Н. Сидоров и И. С. Ильинская <161> еще более решительно исключали из числа безличных предложения типа *По груше упало с дерева; Отцом дано детям по груше*, соотносительные с личными предложениями типа

Груша упала с дерева, Отцом дана детям груша (ср. стр. 153). Р. Ружичка пошел в этом отношении дальше своих предшественников, распространив те же принципы анализа на большинство типов безличных предложений.

Свою основную задачу Р. Ружичка видит в том, чтобы свести каждый тип *б е з л и ч н ы х* предложений с помощью той или иной последовательности обратных трансформаций к определенному типу *л и ч н ы х* ядерных предложений. Рассмотрим без дальнейших комментариев некоторые примеры.

- I. 1. *Листву гонит ветром.* ← *Ветер гонит листву.*
2. *Писем накопилось.* ← *Письма накопились.*
3. *На дворе темно.* ← *Двор темен.*
4. *Светает.* ← *Даль светает* (К. Симонов).
5. *Ноги знобит.* ← *Снег знобит ноги* (Г. Успенский).
6. *Известно о прорыве.* ← *Известен прорыв.*
7. *О конкурсе было объявлено.* ← *Они объявили о конкурсе.*

В этих случаях безличное предложение представляет собой трансформ, сводимый к личному ядерному предложению одной трансформацией.

В ряде случаев «трансформацию безличности можно рассматривать как разновидность пассивной трансформации».

- II. 1. *Документов не сохранилось.* ← *Документы не сохранились.* ← *Документы сохранились.*
2. *Ответа не пришло.* ← *Ответ не пришел.* ← *← Ответ пришел.*

В этих случаях безличное предложение находится на расстоянии двух трансформаций от личного ядерного предложения; трансформацию безличности «можно рассматривать как разновидность трансформации отрицания».

- III. 1. *Деформаций не обнаружено.* ← *Деформации не обнаружены.* ← *Деформации обнаружены.* ← *← Некто обнаружил деформации.*
2. *Препятствий не предвиделось.* ← *Препятствия не предвиделись.* ← *Препятствия предвиделись.* ← *Некто предвидел препятствия.*

Безличное предложение отдалено от личного ядерного предложения на три трансформации; трансформации отрицания предшествует пассивизация.

Подвергнув систематическому трансформационному анализу большинство типов безличных предложений, Р. Ружичка приходит к выводу, что «единство безличных предложений в трансформационной грамматике исчезает. Они входят по своей структуре в разные типы небезличных предложений... Область производных безличных ядерных предложений ограничивается классом слов категории состояния, не входящих в парадигму прилагательных» <159, 31>, ср. *надо, нельзя, жаль*.

* * *

На этом мы заканчиваем рассмотрение моделей исследования языка. Как мы попытались показать, модели данного типа вырабатывают информацию о единицах языка, классах единиц и правилах их связи. Они не дают ответа на вопрос о том, каким образом язык функционирует. Пользуясь образом А. Мартине <334>, мы могли бы сказать, что исследовательские модели дают нам представление об анатомии языка, но ничего или почти ничего не говорят о его физиологии, хотя именно физиология языка представляет наибольший интерес для лингвиста. «Анатомия свелась бы к разновидности «некрологии», или науки о трупах, бесполезной и не представляющей интереса ни для кого, кроме, может быть, профессиональных бальзамировщиков, если бы ее изучение не имело целью объяснить физиологию» <334, 125>. Точно так же и лингвистические понятия, выработанные в процессе «анатомического» исследования, обретают смысл и вес лишь в том случае, когда они используются для объяснения речевого поведения человека. Здесь, однако, возникает новый объект и вступают в силу новые модели — модели речевой деятельности. Им и будет посвящена следующая часть нашей книги.

МОДЕЛИ РЕЧЕВОЙ ДЕЯТЕЛЬНОСТИ

Данная часть посвящена основному типу лингвистических моделей — моделям речевой деятельности человека. Достигнутый в этой области прогресс весьма значителен. Разработаны первые варианты моделей фонологического анализа и синтеза <235>, <236>, <315>, разлагающие звуковой поток в фонологические признаки и синтезирующие важные компоненты речи, включая ее интонационные характеристики <315>. Для многих языков построены алгоритмы автоматического морфологического анализа <116>, <126>, <114>, которые каждую словоформу разлагают на ее основу и информацию о выражаемых словоформой грамматических значениях (например, значениях падежа, числа и т. п. для существительных, значениях лица, времени, наклонения и т. п. для глаголов и т. д.). Имеются алгоритмы автоматического синтеза, которые конструируют некоторую словоформу по ее основе и набору информации о грамматических значениях этой словоформы <40>, <194>, <62 а>. Наиболее систематическим изложением морфологии с указанной здесь синтетической точки зрения, восходящей к упоминавшимся в I части работам Ф. Брюно и О. Есперсена, является докторская диссертация А. А. Зализняка «Классификация и синтез именных парадигм современного русского языка» <62 а> ¹. Существует большое число синтаксических алгоритмов анализа и синтеза текстов <114>, и начинается разработка семантических моделей анализа и синтеза <97>, <108>, <109>, <110 а>. Имеется, наконец, несколько классов порождающих моделей, имитирующих

¹ Она была защищена после окончания данной книги и поэтому не могла быть учтена в ней должным образом.

способность человека отличать правильное от неправильного в языке и строить правильные языковые объекты.

Поскольку все упомянутые здесь типы моделей мы рассмотреть не сможем, необходимо выбрать те из них, на материале которых можно достаточно полно изложить круг методов и проблем, связанных с этой областью современной структурной лингвистики. К числу таких моделей относятся синтаксические и семантические¹.

Глава I

ПОРОЖДАЮЩИЕ МОДЕЛИ

Как помнит читатель, порождающей моделью, излагаемой обычно в форме исчисления, называется конечный набор правил, способных задать или породить все правильные, и только правильные, объекты некоторого множества, в том числе бесконечного, и приписать каждому объекту определенный анализ. Мы рассмотрим три типа синтаксических порождающих моделей: модель порождения по непосредственно составляющим, трансформационную порождающую модель и аппликативную модель.

МОДЕЛЬ ПОРОЖДЕНИЯ ПО НС

В предыдущей части мы говорили о том, как правила свертывания по НС, получаемые в синтаксической модели исследования, могут быть использованы для анализа предложений (свертывания по НС). Оказывается, что те же самые правила, но только обращенные и применяемые

¹ Отметим, что, если классической грамматике свойствен известен морфологизм, сложившийся, вероятно, под влиянием сравнительно-исторического языкознания, которому наиболее богатые материалы поставляли фонетика и морфология, для современной структурной лингвистики характерна ориентация на синтаксис и семантику; семантические и синтаксические механизмы признаются, таким образом, основными механизмами языка. Интересно, что центральное место синтаксиса в грамматической теории понимали наиболее дальновидные представители классического языкознания (см. А. А. Шахматов <212, 195>, А. М. Пешковский <141, 52>, В. В. Виноградов <30, 29>).

в обратном порядке, могут быть использованы и для порождения предложений (развертывания по НС), ср.

Правила свертывания Правила развертывания

- 1) $D + A_x \rightarrow A_x$
- 2) $A_x + \tilde{N}_x \rightarrow \tilde{N}_x$
- 3) $V + \tilde{N}_x \rightarrow V$
- 4) $N_n + \tilde{V} \rightarrow S$

- 1) $S \rightarrow N_n + V$
- 2) $V \rightarrow V + N_x$
- 3) $N_x \rightarrow A_x + N_x$
- 4) $A_x \rightarrow D + A_x$

По этим правилам может быть порождено, в частности, предложение *Мой друг читает очень интересную книгу* (помимо многих других предложений с той же НС-структурой; см. рис. 8).

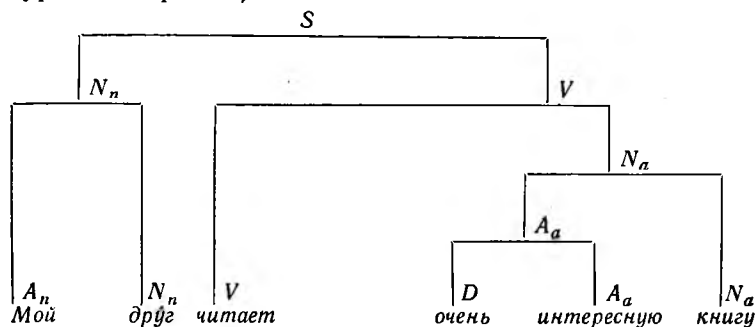


Рис. 8.

Варианты таких правил фактически используются в большинстве разрабатываемых и уже действующих моделей порождения по НС. Наиболее интересной из них является модель, принадлежащая В. Ингве <381>, <382>, <383>, <152>. На ней мы и сосредоточим внимание в дальнейшем.

Модель состоит из 1) грамматики и 2) механизма. Грамматика — это конечное неупорядоченное множество правил развертывания по НС следующих типов: а) $A \rightarrow B + C$ (символ в левой части правила развертывается в два новых символа), например $S \rightarrow NP + VP$, где NP — группа подлежащего, а VP — группа сказуемого; $NP \rightarrow T + N$, где T — артикль; $VP \rightarrow V + NP$; $VP \rightarrow V + Cl$, где Cl — придаточное предложение; $Cl \rightarrow C + S$, где C — союз, а S — предложение (рекурсивный элемент) и т. д.; б) $A \rightarrow B$ (символ в левой части правила заменяется одним, обычно терминальным символом, т. е. символом,

к которому не применимы никакие другие правила грамматики НС), например: $T \rightarrow the, a$ (определенный или неопределенный артикль); $V \rightarrow see$ — «видеть», $hear$ — «слышать», $know$ — «знать» и т. д. с показателями времени, числа, лица; $N \rightarrow girl$ — «девочка», boy — «мальчик», man — «мужчина» и т. д. с показателями числа; в) третью группу составляют правила типа $B \rightarrow D + \dots + E$, предназначенные для порождения конструкций с так называемыми разрывными составляющими, т. е. составляющими типа *если...*, *то* или *поскольку...*, *постольку*, между частями которых могут помещаться другие составляющие. Эти правила применяются следующим образом. Пусть B есть НС конструкции $B + C$, полученной по правилу $A \rightarrow B + C$; тогда результатом применения правила $B \rightarrow D + \dots + E$ должна быть цепочка $D + C + E$ (B — разрывный конститuent).



Рис. 9.

Как видим, правила НС, принятые в предложенной В. Ингве модели, отличаются в некоторых отношениях от обычных правил НС: во-первых, они заранее никак не упорядочены; во-вторых, они усовершенствованы таким образом, что становится возможным порождение конструкций с разрывными составляющими.

Второй частью модели является механизм — идеализированная, но физически реализуемая вычислительная машина, состоящая из четырех взаимосвязанных частей (рис. 9).

Устройство вывода печатает по одному выводимые элементы. В решающем устройстве помещается один символ, разворачиваемый в данный момент по одному из описанных выше правил. В постоянной памяти хранятся все правила грамматики (правила развертывания по НС). В быстродействующей памяти хранятся промежуточные результаты.

Хотя на множестве грамматических правил заранее не установлено никакого отношения порядка, это не значит, что на любом шаге порождающего процесса можно применить любое правило. Выбор каждого следующего правила зависит от того, какое правило было выбрано на предыдущем этапе; первым, по таблице случайных чисел (т. е. случайным образом), выбирается одно из правил развертывания символа $S : S \rightarrow B$, или $S \rightarrow A+B$, или $S \rightarrow C+D$ и т. д.

Чтобы иллюстрировать эти принципы, рассмотрим пример порождения предложений по правилам следующей простой грамматики: 1) $N \rightarrow \text{man}$ — «мужчина», boy — «мальчик»; 2) $S \rightarrow NP+VP$; 3) $V \rightarrow \text{heard}$ — «слышал»; saw — «видел»; 4) $NP \rightarrow T+N$; 5) $T \rightarrow \text{the}$ (определенный артикль); 6) $VP \rightarrow V+N$. В начале порождающего процесса символ S засылается в решающее устройство. Решающее устройство делает две вещи: во-первых, оно подает поступивший в него символ в устройство вывода для печати; во-вторых, оно обращается к грамматике, хранящейся в постоянной памяти, чтобы развернуть символ по какому-либо из ее правил. Применено может быть любое правило, левый символ которого, стоящий до стрелки, совпадает с символом, находящимся в данный момент в решающем устройстве. В нашем случае таким правилом является $S \rightarrow NP+VP$. Каждый раз, когда применено какое-то грамматическое правило, крайняя левая непосредственно составляющая полученной в результате конструкции (в нашем случае — составляющая NP конструкции $NP+VP$) направляется в решающее устройство, а все стоящие справа от него символы (в нашем случае — VP) — в быстродействующую память, где они ждут своей очереди на развертывание. Если находящийся в решающем устройстве символ не является терминальным (см. определение на стр. 199), то решающее устройство обрабатывает его по тем же правилам, что и символ S , т. е. подает его в устройство вывода для печати и обращается к грамматичес-

ким правилам, чтобы заменить его новым символом или символами. Если же символ является терминальным, то решающее устройство подает его на выход, обращается к быстродействующей памяти, хранящей промежуточные результаты, и обрабатывает по тем же правилам первый стоящий на очереди символ. Если в быстродействующей памяти никаких промежуточных символов больше нет, то решающее устройство подает на выход точечный знак (конец предложения).

Изложенный алгоритм порождает по правилам нашей грамматики, в частности, следующее предложение: *The man saw the boy* — «Мужчина увидел мальчика». Процесс порождения этого предложения может быть представлен следующей схемой (см. схему 3).

Добавление к этой простой грамматике некоторых рекурсивных правил позволяет порождать предложения сколь угодно большой длины, т. е. бесконечное множество предложений. Среди этих правил имеются три правила порождения предложений с однородными членами, одно правило порождения предложений с последовательным подчинением и одно правило порождения сложных предложений (с рекурсивным элементом *S*).

Все порождаемые предложения обязаны быть «грамматически правильными», но не обязаны быть осмысленными. Приведем образцы порождаемых машиной предложений (основой для них послужили грамматика и словарь, содержащиеся в десяти предложениях книжки для детей, за тем исключением, что грамматические правила были расширены указанным выше образом):

1. *He has four polished sand-domes* — «У него (человека) есть четыре блестящих песчаных купола».

2. *He has four proud, little, polished, polished and proud boilers under proud bells, steam and whistles in four whistles* — «У него есть четыре гордых, маленьких, отполированных, отполированных и гордых котла под гордыми звоночками, пар и свистки в четырех свистках».

3. *The steam makes its steam and a whistle big and proud* — «Пар делает свой пар и свисток большими и гордыми».

4. *When engineer Small is oiled, the water in the bells is heated* — «Когда инженер Смол смазан, вода в звонках подогревается».

Выход	Решающее устройство	Быстродействующая память
	<i>S</i>	
<i>S</i>	<i>NP</i>	<i>VP</i>
<i>S NP</i>	<i>T</i>	<i>N VP</i>
<i>S NP T</i>	<i>the</i>	<i>N VP</i>
<i>S NP T the</i>	<i>N</i>	<i>VP</i>
<i>S NP T the N</i>	<i>man</i>	<i>VP</i>
<i>S NP T the N man</i>	<i>VP</i>	
<i>S NP T the N man VP</i>	<i>V</i>	<i>NP</i>
<i>S NP T the N man VP V</i>	<i>saw</i>	<i>NP</i>
<i>S NP T the N man VP V saw</i>	<i>NP</i>	
<i>S NP T the N man VP V saw NP</i>	<i>T</i>	<i>N</i>
<i>S NP T the N man VP V saw NP T</i>	<i>the</i>	<i>N</i>
<i>S NP T the N man VP V saw NP T the</i>	<i>N</i>	
<i>S NP T the N man VP V saw NP T the N</i>	<i>boy</i>	
<i>S NP T the N man VP V saw NP T the N boy</i>		

5. *Engineer Small is polished* — «Инженер Смол отполирован».

6. *A fire-box is proud of Small* — «Топка гордится Смолом».

В других экспериментах порождались предложения и более сложной синтаксической структуры <336>:

1. *After he is covered, it never admires his steam* — «После того как он покрыт, оно никогда не восхищается его паром».

2. *In what way isn't a rug old, and why is he wiry and cool* — «В каком смысле коврик не является старым и почему он жилист и хладнокровен».

3. *She makes six boilers and six tragic and new legs* — «Она делает шесть котлов и шесть трагических и новых ног».

4. *What is he cooled for* — «Зачем он охлаждается?»

5. *It no longer runs it, nor does she put it under four smooth floors* — «Оно больше им не управляет, и не кладет она его под четыре гладких пола».

6. *Isn't she polished under Miss Macpherson and plants* — «Ну не отполирована ли она под Мисс Макферсон и растениями».

Как видим, машина довольно успешно учится говорить на человеческом языке, и свои первые шаги на этом новом и трудном для нее поприще она делает в некоторых отношениях гораздо уверенней, чем ребенок, осваивающий свой родной язык, или взрослый, изучающий иностранный язык. Заметим, что для порождения приведенных выше предложений машине потребовалось знание весьма тонких грамматических механизмов.

По мнению В. Ингве, программа, использованная для порождения этих предложений, не обнаружила никаких существенных недостатков. «До сих пор не было причин сомневаться в том, что можно будет использовать ту же самую формальную систему в полной грамматике английского языка» <382, 8>.

Аналогичные принципы положены в основу грамматик НС французского, арабского, русского и других языков, разрабатываемых группой В. Ингве в США и его последователями в других странах. Ниже мы остановимся только на одной из таких грамматик. Это — грамматика порождения по НС для русского языка, предложенная Н. Г. Ар-

сентяевой и представляющая собой в некоторых отношениях существенный шаг вперед <10>, <11>.

Отличие модели Н. Г. Арсентьевой от рассмотренной выше модели В. Ингве состоит в том, что заложенный в модель словарь разбит на семантические классы (два класса — переходных и непереходных — глаголов, причем для каждого глагола указан тип управления, 33 класса существительных, 29 классов прилагательных, 7 классов наречий и т. д.). Для каждого конкретного глагола указывается, из каких классов существительных он может принимать подлежащее и дополнения, а также из каких классов наречий — обстоятельства. Для каждого существительного указаны аналогичным образом классы прилагательных и причастий и т. п. Это обеспечивает порождение по более или менее обычным правилам развертывания по НС частично осмысленных предложений; именно, в порождаемых предложениях имеется осмысленность внутри элементарной синтагмы, хотя осмысленной связи между синтагмами еще нет.

Правила грамматики (свыше 200) и словарь, заложенные в машину, были получены в результате произведенного человеком анализа 50 предложений из «Повестей Белкина»; каждому правилу была приписана, в результате того же анализа, некоторая вероятность выбора (об этом понятии см. стр. 117). Модель была запрограммирована; машинные эксперименты дали следующие результаты (приводятся образцы порожденных предложений, общее число которых очень велико):

1. *Вы будете беспокоить ваших стариков.*
2. *В самом деле я не ошибаюсь.*
3. *В трех верстах от лесов ваших вы ожидаете лошадь вашу долго.*
4. *Вы наедине делали весь ром, находившийся за переписками стройных красавиц в трех верстах от богатых и счастливых домиков.*
5. *Мы, заметя старого или решительного сына, найдем лошадь твоих прапорщиков.*
6. *Ты в год для твоих картинок или романов для особы, матю, находившейся под мешком жителя, достопамятным вашей жене, на красных деньгах лошадей наших стройных молодцов избранной, будешь пылать под пистолетами или пистолетом, стройными, стройными невестами избранными.*

Для сравнения приведем образцы предложений, порожденных первыми вариантами модели, в которых словарь не был разбит на семантические классы.

1. *Разумеется, что бедным или нечаянным молодцам подадут уединение.*

2. *Накануне следующей склонности уездного мешка милый сосед, крестный жаркий день накануне мучительного или мучительного уединения.*

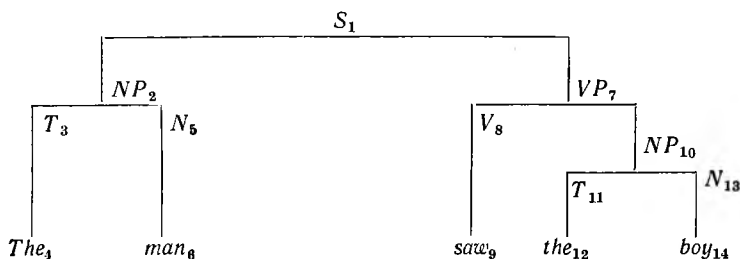
3. *Будут прочить объяснения своих за жителя нашего сосновые воображения.*

4. *Отставная отставная станция, красота постарше, чем ты, будет думать о воскресном нечаянном чтении в трех верстах от крестных армейских голосов.*

5. *Накануне отставного богатого объяснения мы жили, когда я поспешу, может быть, до меня.*

6. *Клянись нам в свиданиях на пронзительных или тех же воображениях ты.*

Чтобы оценить степень адекватности рассматриваемых здесь порождающих грамматик структуре естественных языков, проанализируем их более внимательно. Представим процесс порождения предложения *The man saw the boy* в виде дерева и пронумеруем все имеющиеся в нем узлы, включая терминальные, в порядке их развертывания:



Легко заметить, что машина сначала обрабатывает крайнюю левую ветвь дерева; дойдя до конца этой ветви (точки The_4), она возвращается к ближайшему узлу, в котором началось ветвление влево (NP_2), и заканчивает крайний левый куст дерева. Затем она возвращается к следующему узлу дерева (S_1) и снова обрабатывает левую ветвь очередного куста и т. д. Поскольку любая левая ветвь заставляет машину возвращаться к некоторой исходной точке, ветвящиеся влево структуры назы-

ваются регрессивными. Ветвящиеся вправо структуры, не обладающие этим свойством, называются соответственно прогрессивными.

Когда машина обрабатывает регрессивную структуру, она должна хранить в быстросействующей памяти все промежуточные результаты (символы), причем чем дальше влево ветвится структура, тем больше число символов, которые необходимо запомнить. Для дерева, изображенного на рисунке 10, I, машина должна хранить в памяти максимум два символа, а для дерева, изображенного на рисунке 10, II, число таких символов равно трем.

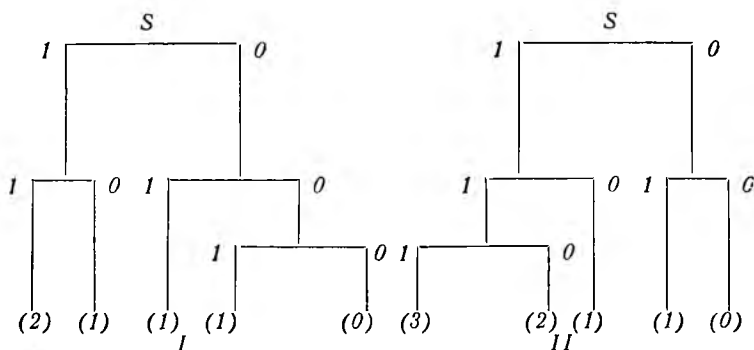


Рис. 10.

Последнее дерево соответствует, например, предложению *Хорошо переведенный текст читается легко*. Первым действует правило $S \rightarrow NP + VP$; VP направляется в быстросействующую память, а NP разворачивается по правилу $NP \rightarrow A + N$; N в свою очередь направляется в быстросействующую память, а A разворачивается по правилу $A \rightarrow D + A$; наконец, A отправляется в быстросействующую память, которая теперь содержит три символа (VP , N и A), а D заменяется символом *хорошо*.

Максимальное число символов, которое машина должна хранить в быстросействующей памяти при порождении некоторого предложения, называется *глубиной* данного предложения. Глубина максимально глубокого предложения называется *глубиной* данного языка. Глубину предложения можно подсчитать, пронумеровав ветви справа налево в точках ветвления, как показано на рисунке 10, и сложив единички, встречаю-

щиеся на пути к терминальной точке (см. цифры в круглых скобках). Легко заметить, что глубина предложений и, следовательно, языков растет исключительно за счет регрессивных структур, так как при развертывании прогрессивных структур машина не должна запоминать значительных по количеству промежуточных результатов.

Из сказанного следует, что модель В. Ингве может порождать предложения с прогрессивными структурами сколь угодно большой длины. Она, однако, не может порождать предложений с бесконечно растущими регрессивными структурами, так как это потребовало бы быстродействующей памяти неограниченного (бесконечного) объема, а у всякой практически реализуемой машины объем памяти обязан быть конечным. Таким образом, глубина порождаемых машиной предложений ограничена максимальным объемом ее быстродействующей памяти.

После этих предварительных замечаний мы можем перейти к наиболее интересным идеям, связанным с моделью порождения по НС.

Каждая грамматическая модель неизбежно является некоторой гипотезой о том, как устроен в действительности механизм, работу которого она имитирует. Если подходить к лингвистическим моделям с этой точки зрения, то каждому механизму модели следует поставить в соответствие некий психический или физиологический механизм. При этом следует помнить, что механизмы модели всегда проще, чем соответствующие им механизмы мозга, и являются в лучшем случае аппроксимацией последних (см. стр. 81).

Порождающий механизм рассмотренной выше модели В. Ингве состоит, как мы помним, из четырех частей. Постоянной памяти естественно поставить в соответствие память, а устройству вывода — речевые органы. Если мы хотим быть последовательными, мы должны предположить существование в мозге человека и других двух механизмов, соответствующих быстродействующей памяти и решающему устройству. Оказывается, что в пользу этого предположения можно привести весьма любопытные факты и соображения.

Психологическими экспериментами установлено, что человек в процессе обработки информации может одновременно хранить в памяти ограниченное число единиц, а именно 7 ± 2 случайных чисел, не связанных друг с другом

слов и т. п. <339>. Этим объясняется, в частности, тот факт, что ему трудно производить и понимать предложения с большой глубиной типа *If what going to a clearly not very adequately staffed school really means is little appreciated we should be concerned* <381> — «Если (в обществе) плохо понимают, что в действительности значит ходить в школу, штат которой явным образом не очень хорошо укомплектован, это должно нас озаботить» (глубина предложения равна 8). Здесь человек действует на пределе своих возможностей, так как заполнение его быстродействующей памяти близко к критической отметке (точке 9). Рассмотрим еще следующий пример: (1) *Она знала человека, который сделал сообщение, которое удостоилось премии, которая была назначена за лучшую работу* (ср. <341>). Это предложение, при всех его стилистических изъянах, построено правильно и легко понимается. Будем теперь постепенно увеличивать его глубину, перестраивая его с помощью грамматического механизма в л о ж е н и я предложений; этот механизм состоит в том, что в сложноподчиненном предложении одно из составляющих его простых предложений разрывается на две части, и между ними вкладывается другое простое предложение. Результатом однократного использования вложения является предложение (2) *Человек, которого она знала, сделал сообщение, которое удостоилось премии, которая была назначена за лучшую работу*. Хотя глубина увеличивается, предложение сохраняет удобопонятность. С гораздо большим трудом воспринимаются предложения (3) и в особенности (4) с соответственно возрастающей глубиной: (3) *Сообщение, которое человек, которого она знала, сделал, удостоилось премии, которая была назначена за лучшую работу*; (4) *Премия, которой сообщение, которое человек, которого она знала, сделал, удостоилось, была назначена за лучшую работу*.

Поскольку объем быстродействующей памяти человека ограничен, естественно поставить вопрос о том, оказывает ли это какое-либо влияние на структуру языка. Если такое влияние действительно имеет место и проявляется в том, что язык вырабатывает грамматические механизмы, ограничивающие глубину предложений, то изложенные выше модели порождения по НС адекватны структуре естественных языков; если же указанные выше особенности строения быстродействующей памяти не оказывают никакого влияния на структуру языка, т. е. никак

не ограничивают грамматически допустимую глубину предложений, то модель порождения по НС, в которой такие ограничения неизбежны, не вполне адекватна структуре естественных языков.

В. Ингве полагает, что такое влияние имеет место и проявляется в действии ряда механизмов, к числу которых он относит:

1) Предпочтение, отдаваемое языком постпозиции по сравнению с препозицией. Группы, находящиеся в постпозиции, в большинстве случаев ветвятся вправо, ср. предложение (1) из приводившегося на стр. 209 примера. Группы, находящиеся в препозиции, в большинстве случаев ветвятся влево и тем самым увеличивают глубину предложения; ср. предложение (4) из того же примера. В. Ингве полагает, что при большой глубине такие предложения утрачивают грамматичность¹.

2) Предпочтение, отдаваемое языком бинарным структурам по сравнению с тернарными; дерево с тремя ветвями (тернарное), изображенное на рисунке 11, I, имеет глубину 2, в то время как соответствующее ему дерево с двумя ветвями и правым втвлением, изображенное на рисунке 11, II, имеет глубину 1.

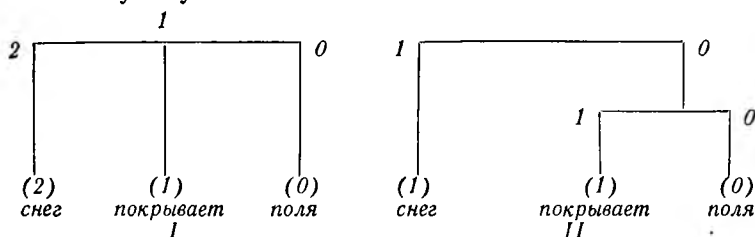
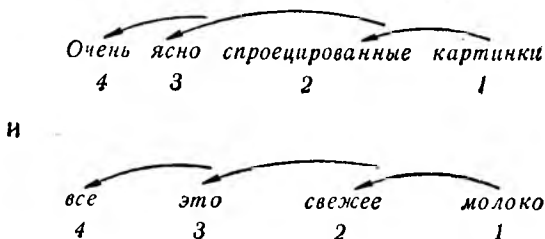


Рис. 11.

3) Механизмы структурной перестройки предложений, глубина которых близка к критической отметке. Вместо глубокой фразы типа *Более высокий, чем цирковой гигант, человек* мы предпочитаем менее глубокую (и более грамматичную) фразу *Человек, более высокий, чем цирковой гигант*. Вместо глубокой фразы *Он увидел шире первого поле* мы предпочитаем менее глубокую фразу *Он увидел поле, которое было шире первого* и т. д.

¹ Ср. интересные статистические данные Г. А. Лескиса (99, 13), которые как будто подтверждают эти соображения.

4) Механизм классов, который ограничивает глубину регрессивных структур определенных типов числом 4—5. К такого рода структурам В. Ингве относит именные словосочетания типа



И в том и в другом случае глубина словосочетания является предельной, так как не существует класса слов, элементы которого могли бы зависеть от наречия типа *очень*, и не существует класса прилагательных, элементы которого могли бы служить определением ко *в с е м у* словосочетанию *Все это свежее молоко* <309, 189>.

Этих и им подобных фактов, однако, недостаточно для действенного подтверждения гипотезы. Не вполне ясно, например, как может выглядеть грамматическое правило, ограничивающее глубину словосочетаний типа *Все это свежее молоко* каким-либо определенным числом (см. обсуждение этого вопроса на стр. 85). Р. Стокуэлл <359, 44> приводит, например, предложение со словосочетанием того же типа, глубина которого, по-видимому, превосходит критическую¹: *Все эти таинственные маленькие древние квадратные черные китайские бумажные коробочки находятся здесь*. Предложения типа *Сообщение, которое человек, которого она знала, сделал, удостоилось премии, которая была назначена за лучшую работу*, хотя и менее привычные, чем соответствующие предложения с меньшей глубиной, являются все-таки правильно построенными. Интересно, что психологические эксперименты <341> показывают колоссальное различие в восприятии таких предложений и действительно неправильно построенных предложений типа **Сделал она человек удостоилась была сообщение за которая премии* и т. д. Следует, наконец, указать и на факт существования языков, в которых регрессивные

¹ При условии, что прилагательные в словосочетаниях такого типа не считаются однородными членами.

структуры являются не менее, если не более, обычными, чем прогрессивные структуры. К числу таких языков относятся, например, японский <250> и венгерский <373 а>.

В свете изложенного становится понятным, почему некоторые исследователи считают модель порождения по НС эффективной лишь в определенных рамках. С их точки зрения, такая модель может быть верной лишь постольку, поскольку она является предположением о возможном устройстве механизма порождения текстов, находящегося в мозге человека <340>; однако она не вполне адекватна структуре языка в целом, поскольку она не может породить некоторых типов предложений, для производства которых у языка есть все необходимые средства. Это и побудило некоторых ученых, прежде всего Н. Хомского, предпринять разработку нового — т р а н с ф о р м а ц и о н н о г о — типа порождающей модели.

ТРАНСФОРМАЦИОННАЯ МОДЕЛЬ ПОРОЖДЕНИЯ¹

Разработка трансформационной порождающей грамматики была начата Н. Хомским свыше 10 лет назад <262>, <199>, <200>, <202>, <265>, <264>, <239>, <240>, <325>, <100>, <101>, <329>, <354>, <359>, <204>. В настоящее время исследованиями в области трансформационной грамматики занимаются большие коллективы ученых разных стран, и интерес к ней не только не падает, как склонны полагать некоторые ученые <197, 109>, <255, 366>, <129, 294>, но год от года возрастает. Были написаны значительные фрагменты трансформационных грамматик английского, немецкого, русского, испанского, хинди, финского, эстонского, турецкого, японского, луганда, арабского, тагальского и других языков <240, 89>. Одним из наиболее серьезных трудов в этой области являются фрагменты трансформационной грамматики немецкого языка, написанные группой молодых немецких ученых <361>, <362>².

¹ В данном разделе излагается один из ранних вариантов трансформационной грамматики. До проработки этого раздела мы рекомендуем читателю восстановить в памяти идеи, связанные с трансформационным методом в целом (см. стр. 181 и след.). Почти все они сохраняют силу применительно к моделям порождения и образуют естественное введение к ним.

² См. также интересные опыты в <14>, <16>, <132>.

Основная идея трансформационной порождающей грамматики заключается в следующем. Язык представляет собой сложную, многоуровневую структуру, причем законы строения единиц высших уровней (например, сложных предложений) в общем случае не совпадают с законами строения единиц низших уровней (например, слов). Так, порядок морфем внутри слова в большинстве языков достаточно строго фиксирован (*бел-изн-а*, *черн-от-а* — правильные формы, а *изн-бел-а*, *от-черн-а* — неправильные), в то время как внутри предложения такие перестановочные трансформации возможны (ср. *пришел мальчик* — *мальчик пришел*). Законы строения единиц высших уровней не только отличаются от законов строения единиц низших уровней, но и невыводимы из них; поэтому единицы более высоких уровней не являются простой комбинацией единиц непосредственно предшествующих уровней. Наиболее ярким проявлением указанных здесь свойств считается тот факт, что некоторые процессы и явления данного уровня, например омонимию, нельзя описать, пользуясь понятиями предшествующего уровня. Как свидетельствует Р. И. Аванесов <1, 52>, фразы *и скота* и *из кота*, равно как и фразы *и скита* и *из кита*, произносятся одинаково: [и/ск˘тá], [и/ск'итá]. Чтобы различить омонимию, т. е. показать, что фразы [и/ск˘тá], [и/ск'итá] могут быть проанализированы двумя различными способами и имеют, следовательно, два различных значения, необходимо обратиться к понятиям морфологического уровня. Морфологический анализ приводит нас к заключению, что фразы отличаются друг от друга составом морфем; с этим и связано различие в их значениях. Возьмем теперь фразу *Старые мужчины и женщины*. Для нее морфологические соображения теряют силу, так как при неизменном составе морфем она может быть понята либо в смысле *Старики и старухи (старые мужчины и старые женщины)*, либо в смысле *Старики и женщины (старые мужчины и всякие женщины)*. Для того чтобы различить омонимию в этой фразе, мы должны перейти с морфологического на более высокий — синтаксический — уровень, прибегнув, как минимум, к анализу по НС, ср. ((*старые мужчины*) и *женщины*) или (*старые (мужчины и женщины)*). Для анализа фразы *Сплочение рабочих бригад*, содержащей, как мы помним, четыре омонима и допускающей, следовательно, четыре различных осмысления, не

годится даже анализ по НС, так как структура (дерево) НС для этой фразы во всех четырех случаях одна и та же. Для решения омонимии необходимо привлечь более сильные соображения следующего — трансформационного — уровня. Наконец, для анализа фразы *Он пошел в институт* недостаточны, по-видимому, и трансформационные соображения, и необходим чисто семантический анализ: 1) «Он отправился в институт (а не в магазин)»; 2) «Он поступил в институт (а не в армию)».

По обеим указанным причинам адекватная грамматическая модель должна содержать несколько уровней описания, соответствующих уровням структуры языка, причем каждый уровень описания должен представлять собой специфический набор правил.

Трансформационная порождающая грамматика содержит три уровня описания, каждый из которых задается частично упорядоченным множеством правил определенного вида и является одним из компонентов трансформационной порождающей грамматики. К числу таких компонентов относятся: 1) правила НС, 2) трансформационные правила и 3) морфонологические правила¹. Ниже мы рассмотрим более подробно первые две части, оставив в стороне морфонологические правила, т. е. обязательные правила подстановки определенных алломорфов вместо морфем конечной цепочки порождающего процесса.

Исходными для трансформационной порождающей грамматики являются правила развертывания по НС, несколько отличные от тех, которые рассматривались нами выше. Так как в трансформационной грамматике на их долю приходится лишь незначительная часть работы, выполняемой в процессе построения предложения, появляется возможность значительно уменьшить число правил НС и упростить их структуру.

Правила развертывания по НС в трансформационной порождающей грамматике суть правила подстановки вида $A \rightarrow B + C$, на которые наложены следующие ограничения:

1) Каждое правило применяется к цепочке определенного внешнего вида, совершенно независимо от того, каким образом эта цепочка была получена. Так, одни и те же правила развертывания символа NP (группы суще-

¹ В последнее время в трансформационную грамматику встраивается четвертый — семантический — компонент <325а>.

ствительного) применимы и в том случае, когда NP стоит до VP (группы глагола), т. е. получен по правилу $S \rightarrow NP + VP$, и в том случае, когда NP стоит после V , т. е. получен по правилу $VP \rightarrow V + NP$. Единственным необходимым условием применения данного правила является наличие в некоторой цепочке НС символа, который стоит в левой части данного правила. Иными словами, правило $VP \rightarrow V + NP$ можно применить к цепочке вида $NP + VP$, поскольку она содержит символ VP , стоящий в левой части данного правила, но нельзя применить к цепочке вида S .

2) Результатом применения правил НС к исходной цепочке должно быть **дерево** составляющих, так как трансформации применяются именно к деревьям, а не к цепочкам, причем дерево составляющих должно быть приписано данному предложению вполне **однозначным** образом. Поэтому в НС-компоненте трансформационной грамматики запрещается разворачивать более одного символа одновременно. Если бы этого ограничения не было и, допустим, в цепочке $A + B$ можно было бы заменить оба символа сразу с результатом $C + D + E$, то результирующей цепочке нельзя было бы однозначным образом приписать **дерева** НС. Действительно, в данном случае оказываются возможными минимум четыре анализа <240>, <359>.

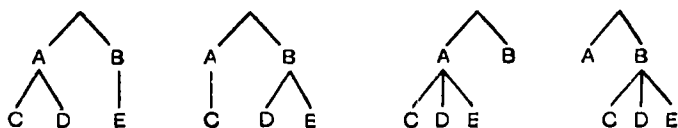


Рис. 12.

3) Запрещены любые рекурсивные правила, то есть правила вида $A \rightarrow A$ или $A \rightarrow A + B$, так как в грамматике НС они часто дают антиинтуитивные решения. В частности, предложениям с однородными членами они приписывают тот же анализ, что и предложениям без однородных членов; ср., например, цепочки *очень очень согретое молоко* и *очень сильно согретое молоко*, которые в этом случае должны иметь дерево одного вида (см. рис. 13), хотя более естественно считать, что таким образом построено только второе словосочетание (*очень сильно согретое молоко*). Этого, конечно, можно было бы избежать следующим образом. Рекурсивные правила вида $A \rightarrow A + B$

сохраняются для порождения предложений с последовательным подчинением типа (((*очень хорошо*) *спроецированные*) *картинки*) (*появились на экране*)), а для порождения предложений с однородными членами вводятся нерекурсивные правила вида $A \rightarrow D + D + D + Adj$, где одно-

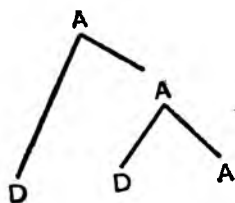


Рис. 13.

родный элемент D может повторяться n раз. Однако таким образом мы лишаем себя преимуществ рекурсивных правил и должны иметь отдельное правило разворачивания цепочки с однородными членами для каждого $n \geq 2$, например $A \rightarrow D + D + Adj$, $A \rightarrow D + D + D + Adj$, $A \rightarrow D + D + D + D + Adj$ и т. д. При этом, чтобы сделать грамматику конечной,

мы должны ограничиться каким-то определенным n , например $n=10$ или $n=100$. В результате грамматика становится неадекватной языку, потому что в естественном языке всегда можно построить предложение, которое будет содержать $n+1$ однородных членов при любом конечном n . Выход из этих трудностей только один — перенести все рекурсивные правила в трансформационный компонент грамматики.

4) Из сказанного следует, что из НС-компонента грамматики должны быть изъяты и правила вида $A + B \rightarrow B + A$ (п е р е с т а н о в к и), которые могут быть истолкованы как две рекурсии. Эти правила также переносятся в трансформационный компонент грамматики.

Правила разворачивания по НС делятся на контекстуально свободные и контекстуально связанные. Первые применяются всегда, а вторые — только в том случае, когда левый символ правила стоит в определенном контексте.

Приведем образцы возможных (но, конечно, не единственно возможных) правил НС <200>:

$$1. S \rightarrow NP + VP$$

$$2. VP \rightarrow Aux \left\{ \begin{array}{l} be \left\{ \begin{array}{l} Pred \\ Adv_1 \end{array} \right\} (Adv) \\ VP_1 \end{array} \right\}^1$$

¹ Фигурная скобка соответствует союзу «либо... либо». Круглые скобки обозначают факультативность. *Adv* значит наречие,

3. $VP_1 \rightarrow V \left(\left\{ \begin{array}{c} NP \\ Pred \end{array} \right\} \right)$
4. $V \rightarrow \left\{ \begin{array}{l} \left\{ \begin{array}{c} V_s \\ becomes \end{array} \right\} \text{ в контекстах } Pred \\ V_t \text{ в контекстах } NP \\ V_i \text{ в контекстах } \left\{ \begin{array}{c} \# \\ Adv \end{array} \right\} \end{array} \right\}$
5. $NP \rightarrow \left\{ \begin{array}{c} NP_{sing} \\ NP_{pl} \end{array} \right\}$
6. $NP_{sing} \rightarrow T + N + \emptyset$
7. $NP_{pl} \rightarrow T + N + S$
8. $V_t \rightarrow \left\{ \begin{array}{l} V_{t_1} \text{ в контексте } N_h \\ \left\{ \begin{array}{c} V_{t_{s1}} \\ V_{t_{s2}} \end{array} \right\} \text{ в контексте } N_h \dots \dots N_c \end{array} \right\}$
9. $T \rightarrow a, the$
10. $Aux \rightarrow C(M) (have + en) (be + ing)$
11. $C \rightarrow \left\{ \begin{array}{c} Past \\ Present \end{array} \right\}$
12. $N_h \rightarrow man — „человек“, boy — „мальчик“, I — „я“ \dots$
 $N_c \rightarrow table — „стол“, book — „книга“, it — „это“ \dots$
 $N_{ab} \rightarrow sincerity — „искренность“, justice — „справед-
ливость“, it — „это“ \dots$
13. $V_i \rightarrow arrive — „прибывать“, disappear — „исчезать“ \dots$
14. $V_s \rightarrow feel — „чувствовать“, seem — „казаться“ \dots$
15. $V_{t_1} \rightarrow admire — „восхищаться“, find — „находить“ \dots$
 $V_{t_2} \rightarrow terrify — „ужасать“, astonish — „удивлять“ \dots$
 $V_{t_{s1}} \rightarrow find — „находить“, complete — „заканчивать“ \dots$
 $V_{t_{s2}} \rightarrow eat — „есть“, smoke — „курить“ \dots$

.....

Выше было сказано, что множество правил развертывания по НС частично упорядочено. Отношение частичного порядка на этом множестве задается первым условием

adj — прилагательное, *M* — модальные глаголы, *Pred* — предикатив, *pri* — предложное наречие, *sing* — единственное число, *pl* — множественное число, \emptyset — флексия существительного ед. числа, *S* — флексия существительного мн. числа, *past* — прошедшее время, *present* — настоящее время, $\#$ — граница. Правила чтения других знаков см. в тексте самих правил.

применения правила НС к цепочке, сформулированным на стр. 215.

Конечным результатом применения правил НС к исходной цепочке является основа ядерного предложения (можно было бы дать следующее вполне строгое определение этого термина: основой ядерного предложения называется конечный результат применения некоторой совокупности правил НС к исходной цепочке S). Так, применение правил 1, 2, 3, 4, 5 (дважды), 6, 7, 8, 9 (дважды), 10, 11, 12 (дважды) и 15 дает, например, цепочку символов *The man ø Past have en find the boy S*, которая является основой предложения *The man had found the boys* (об основе и парадигме предложения см. стр. 185). Формально основы ядерных предложений рассматриваются как аксиомы некоторого синтаксического исчисления $\langle 199 \rangle$, $\langle 14 \rangle$, а содержательно — как языковые аналогичные элементарных ситуаций (см. стр. 184).

На определенном шаге порождающего процесса мы получаем так называемые ядерные типы, дальнейшее развертывание которых и приводит нас к основам ядерных предложений. Система ядерных типов для английского языка в некоторых грамматиках $\langle 359 \rangle$ выглядит следующим образом (возможны, разумеется, и другие системы):

1. $NP + aux + V_i (adv)$, например,
John ran (home) — «Джон бежал (домой)»
The ice fell (on the floor) — «Лед упал (на пол)»
2. $NP + aux + V_t + NP (adv)$, например,
John killed Mary — «Джон убил Мэри»
The heat melted the ice easily — «Жар легко растопил лед»
3. $NP + aux + be + NP (adv)$, например,
John is a bore — «Джон — зануда»
4. $NP + aux + be + adj (adv)$, например,
The problem was difficult — «Проблема была трудна»
5. $NP + aux + V_t + NP + NP (adv)$, например,
They built John an office — «Они построили Джону контору»

6. $NP + aux + V_t + prt + NP$ (*adv*), например,
They took over the country — «Они захватили страну»
7. $NP + aux + V_s + adj$ (*adv*), например,
He felt good — «Он чувствовал себя хорошо».

Этот список показывает в ясной форме, что правила порождения по НС обеспечивают порождение конечного множества предложений 7—10 простейших типов. Основная работа по выведению предложений, начиная с элементарно осложненных и кончая сложнейшими, выпадает на долю трансформационных правил, т. е. второго компонента трансформационной грамматики. В этом заключается другая важная идея трансформационного исчисления, уже упоминавшаяся раньше: небольшое число простых правил, рекурсивно применяемых к небольшому числу простых синтаксических типов, обеспечивает порождение бесконечного множества предложений всех продуктивных типов, независимо от степени их сложности.

Рассмотрим некоторые важнейшие свойства трансформационных правил. Во-первых, как ясно из сделанных выше замечаний, трансформационные правила рекурсивны. Во-вторых, в них можно заменять одновременно более одного символа. В-третьих, в них разрешены перестановки. В-четвертых, они применяются не к цепочкам определенного внешнего вида, а к деревьям предложений; для того чтобы правильно применить ту или иную трансформацию, необходимо знать не только состав цепочки, но и историю ее получения из исходной цепочки, или *д е р и в а ц и о н н у ю и с т о р и ю*. Рассмотрим предложения *Он видит отца семейства* и *Он лишает дочь наследства*. Внешне они не отличаются друг от друга (обоим соответствует цепочка символов $N_n V N_a N_g$). Однако деревья НС для этих предложений различны (см. рис. 14).

Пассивная трансформация, правым трансформом которой являются фразы вида *Дочь лишена им наследства*, возможна для предложений с деревом НС, показанным на рисунке 14, II, и невозможна для предложений с деревом НС, показанным на рисунке 14, I.

а символ ИС — «изменение структуры»; трансформации приводятся по <200>):

1. Пассив. CP: $NP, Aux, V_t, NP (Adv)$
ИС: $X_1 - X_2 - X_3 - X_4 - X_5 \rightarrow X_4 - X_2 + be + en - X_3 - by + X_1 - X_5$
- 2.* Число. (a) CP: $NP_{sing.}, Present, X$
ИС: $X_1 - X_2 - X_3 \rightarrow X_1 - S - X_3$
(b) CP: $NP_{pl.}, Present, X$
ИС: $X_1 - X_2 - X_3 \rightarrow X_1 - \emptyset - X_3$
3. Вопросы-
тельное
предло-
жение. CP: (a) NP, C, VP_1
(b) $NP, C + M, X$
(c) $NP, C + have, X$
(d) $NP, C + be, X$
ИС: $X_1 - X_2 - X_3 \rightarrow X_2 - X_1 - X_3$
4. Отрицание. CP: см. 3.
ИС: $X_1 - X_2 - X_3 \rightarrow X_1 - X_2 + n't - X_3$
5. Номина-
лизаци¹. (a) CP: $\left\{ \begin{array}{l} T, it + \emptyset, VP \\ NP, C, VP_1 \end{array} \right\} = \left\{ \begin{array}{l} X_1 - X_2 - X_3 \\ X_4 - X_5 - X_6 \end{array} \right\}$
ИС: $X_4 + S$, или \emptyset , заменяет X_1 ; $ing + X_6$ заме-
няет X_2
(b) CP: см. (a)
ИС: $to + X_6$ заменяет X_2
(c) CP: так же, как (a), с V_i вместо VP_1
 $ing + X_6 + of + X_4$ заменяет X_2

В качестве иллюстрации покажем вывод предложения *The boys had been found by the man* — «Мальчики были обнаружены мужчиной». Как мы помним, вывод по правилам ИС дает нам на последнем шаге цепочку *the man \emptyset past have en find the boy S*. В этой цепочке *the man \emptyset* соответствует X_1 , *past have en* — X_2 , *find* — X_3 , *the boy S* — X_4 ; X_5 в ней не представлен. По правилам пассивной трансформации X_1 и X_4 меняются местами, после X_2 добавляется *be + en*, а перед X_1 ставится предлог *by*. Произведем все эти изменения в нашей цепочке и получим *the man \emptyset — past have en — find — the boy S \rightarrow the boy S — past have en be en — find — by the man \emptyset* . Применяя обязательные трансформации и морфонологические правила, получаем *The boys had been found by the man*.

¹ Продуктивность и относительно высокая употребительность этого класса трансформаций составляют типологическую особенность английского языка по сравнению с большинством славянских <267, 24> и некоторыми романскими.

Отметим, что практически любой реально встретившийся или потенциально возможный тип предложения описывается как результат применения некоторых трансформаций к предложениям ядерного вида. Так, весьма интересное по структуре предложение *I'm a great believer in no one knowing more than he has to know to do his job* — «Я глубоко верю в то, что никто не должен знать больше, чем ему положено, чтобы справляться со своей работой»¹, отмеченное нами у Ч. Сноу, естественно представляется как результат 11 регулярных трансформаций (не считая обязательных), которым подвергаются пять предложений ядерного вида:

- (1) *I believe in smth greatly.* — «Я глубоко верю в нечто».
- (2) *One knows more.* — «Некто знает больше».
- (3) *He knows smth.* — «Он знает нечто».
- (4) *He does the job.* — «Он справляется с работой».
- (5) *The job is his.* — «Работа — его».

Предложение (1) номинализируется следующим образом: *I believe in smth greatly* → *I am a great believer in smth*. Предложение (2) претерпевает действие двух трансформаций — отрицания и номинализации, ср. *One knows more* → *No one knows more* → *no one knowing more*. Предложение (3) подвергается модальной трансформации и номинализации, ср. *He knows smth* → *he has got to know smth* → *Than he has got to know*. Затем номинализируется предложение (5): *The job is his* → *his job*. Результат подставляется в предложение (4): *He does the job* → *he does his job*. Последний трансформ номинализируется с помощью инфинитивной трансформации: *He does his job* → *to do his job*. Номинализация *No one knowing more* подставляется в трансформ первого предложения: *I am a great believer in smth* → *I am a great believer in no one knowing more*, и к результату последовательно присоединяются трансформы *than he has got to know* и *to do his job*: *I am a great believer in no one knowing more + than he has got to know* → *I am a great believer in no one knowing more than he has got to know + to do his job* → *I am a great believer in no one knowing more than he has got to know to do his job*.

Сравним теперь трансформационное исчисление в целом с исчислением НС. Отметим прежде всего, что 1) транс-

¹ В этом предложении ярко проявляется отмеченная выше типологическая особенность английского языка, свободно использующего номинализации там, где русский язык предпочитает глагол, ср. <200>, <267>.

формационное исчисление обеспечивает порождение предложений таких типов, которые, по-видимому, принципиально невыводимы по правилам НС; 2) трансформационное выведение некоторых сложных синтаксических типов из простых обладает тем преимуществом, что описывает порождающий процесс и его результат более естественным образом; 3) трансформационное исчисление более «проницательно», так как позволяет установить весьма глубокие связи между синтаксической структурой предложения и его значением, которые правилами НС не фиксируются.

АППЛИКАТИВНАЯ ПОРОЖДАЮЩАЯ МОДЕЛЬ ¹

Аппликативная модель принципиально отличается от рассмотренных выше моделей порождения в двух отношениях: своей логической структурой и формализованной в ней лингвистической концепцией.

Своеобразие логической структуры аппликативной модели состоит в том, что порождающий процесс явным образом представлен в ней как протекающий на *д в у х* существенно различных уровнях — уровне конструкторов и уровне наблюдения. Порождающий процесс начинается с выведения идеальных объектов — конструктивных аналогов слов и предложений, которые на втором этапе порождающего процесса превращаются, с помощью определенных правил интерпретации, в реальные слова и предложения того или иного естественного языка.

Порождающий механизм, продуктом которого являются идеальные объекты, работает независимо от правил интерпретации. Идеальные объекты не содержат в себе информации о грамматических (словообразовательных и словоизменяемых) категориях того или иного конкретного языка, таких, как род, число, падеж, лицо, время, наклонение, вид и пр. Эти и им подобные категории возникают лишь при интерпретации модели. Поэтому либо сам

¹ При изложении разработанной С. К. Шаумяном аппликативной порождающей модели <208>, <209>, <210>, <211> мы будем придерживаться текста <211>. Последний вариант аппликативной модели, опубликованный в книге С. К. Шаумяна «Структурная лингвистика» (изд. «Наука», М., 1965), вышел из печати после завершения работы над рукописью данного очерка.

механизм с имеющимися в нем идеальными словами и правилами образования идеальных предложений, либо порождаемое им множество идеальных объектов может рассматриваться как идеальный язык, допускающий, в частности, использование в качестве языка-посредника при типологических сопоставлениях.

Лингвистическое своеобразие аппликативной модели состоит в следующем. В отличие от модели НС и трансформационной модели, в которых объектами операций являются цепочки или деревья, преобразуемые в новые цепочки или деревья, в аппликативной порождающей модели рассматриваются два рода лингвистических объектов — **к л а с с ы** **с л о в** и **к о м п л е к с ы** **с л о в** — и в соответствии с этим в ней предусмотрено два различных, хотя и связанных друг с другом, порождающих механизма — устройство порождения классов слов и устройство порождения комплексов слов. Первое из них моделирует отношения между единицами на парадигматической оси языка, а второе — отношения между единицами на синтагматической оси языка (ср. сходные идеи у И. И. Ревзина <152>).

Лингвистически аппликативная порождающая модель отличается от трансформационной и НС модели еще в одном отношении. В ней используются две различные операции — аппликация, давшая название модели (см. ниже), и трансформация. Аппликация является единственным правилом образования объектов, а трансформация — единственным правилом их инвариантного преобразования.

Рассмотрим подробнее аппарат аппликативной порождающей грамматики. **А л ф а в и т** **с и м в о л о в**, используемых в этой модели, состоит из символов четырех классов идеальных основ: *N* (интерпретируется как класс основ непроизводных существительных типа *дом*, *рыба*, *окно*), *V* (интерпретируется как класс основ непроизводных глаголов типа *идти*, *ехать*), *A* (интерпретируется как класс основ непроизводных прилагательных типа *белый*, *большой*) и *D* (интерпретируется как класс основ непроизводных наречий типа *там*, *тогда*, *туда*)¹. В дальнейшем, для простоты формулировок, мы будем называть *N*, *V*, *A*, *D* непроизводными основами слов соответствующих классов.

¹ В последних вариантах модели рассматривается один исходный символ *O*, обозначающий идеальный корень слова.

Из этих символов по правилам аппликации (этимологически «приложения», «приставления») образуются комплексы. Аппликативные правила образования комплексов можно задать в виде следующей таблицы (см. табл. 9).

Таблица 9

Как видно из этой таблички, *V* и *A* апплицируются к *N*, а *D* апплицируется к *A* и *V*. Минус в клетке значит, что аппликация невозможна, а *O* — что аппликация для соответствующей пары не определена¹. Как видно из расположения минусов, аппликация направлена: если *X* апплицируется к *Y*, то неверно, что *Y* апплицируется к *X*. Подчеркнем, что порядок следования в записи *X* и *Y* не имеет значения.

	<i>N</i>	<i>V</i>	<i>A</i>	<i>D</i>
<i>N</i>	<i>O</i>	—	—	<i>O</i>
<i>V</i>	<i>VN</i>	<i>O</i>	<i>O</i>	—
<i>A</i>	<i>AN</i>	<i>O</i>	<i>O</i>	—
<i>D</i>	<i>O</i>	<i>DV</i>	<i>DA</i>	<i>O</i>

Теперь мы можем дать определение элементарных комплексов: элементарными комплексами называются символы *N*, *V*, *A*, *D*, а также любая *n*-ка (пара, тройка, четверка или пятерка) символов, образованная однократным применением любой комбинации правил аппликации к символам *N*, *V*, *A*, *D*. Иными словами, комплексами в этом смысле будут следующие 14 выражений: *N*, *V*, *A*, *D*, *NV*, *VD*, *AN*, *DA*, *DAN*, *ANV*, *NVD*, *DANV*, *ANVD*, *DANVD*. Их можно интерпретировать как основы слов и «основы» словосочетаний и предложений простейших типов, ср. *здесь установилась очень хорошая погода (DANVD)*, *большой дом (AN)*, *хмурый день брезжит (ANV)*, *часы бьют полночь (NVD)* и т. д.

На основе понятия аппликации определяется иррефлексивное, асимметричное и интранзитивное (см. стр. 114 и 115) отношение а п п л и к а т и в н о й д о м и н а ц и и

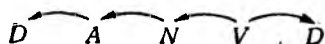
¹ Поясним на примере различие между понятиями невозможности и неопределенности операции. Пусть у нас имеется правило пассивизации (с творительным падежом), применимое к фразам вида $N_n^1 V N_a^2$. Тогда для фраз типа *Он видит меня*, *Книга стоит рубль* пассивизация определена, но невозможна, а для фраз типа *Солнце встает над лесом*, *Петух кричит на заре* пассивизация не определена.

(подчинения по правилам аппликации); это отношение можно задать следующей схемой:



N подчиняет *A* и *V*, *A* и *V* подчиняют *D*. Как видим, с точки зрения аппликативной доминанции центром предложения является существительное (в функции подлежащего); это вполне соответствует традиционной точке зрения на подлежащее как абсолютное определяемое предложения и отражает факт формального согласования сказуемого с подлежащим в роде, числе и лице.

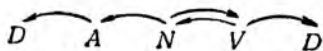
С другой стороны, существует такая точка зрения, с которой центром предложения оказывается глагол (сказуемое). Предложение невозможно без предикативности, нормальным выражением которой является глагол в личной форме; именно он создает предложение (см. стр. 174). Эти и им подобные содержательные соображения формализуются в аппликативной порождающей модели введением еще одного подчинительного отношения — отношения конститутивной доминанции. Оно, разумеется, обладает теми же свойствами, что и аппликативная доминанция, но в одном случае направлено по-другому, не от *N* к *V*, а от *V* к *N*; это дает следующий результат:



Таким образом, для всех связанных отношением подчинения пар, за исключением пары *NV*, направление аппликативной и конститутивной доминанции совпадает, а для пары *NV* оно различно. В итоге мы имеем



или, в упрощенном виде,



¹ Аналогичные идеи об отношениях детерминации и конституирования в предложении и словосочетании высказывал Е. Курилович; см. сноску 2 на стр. 175.

До сих пор мы упоминали лишь одно свойство каждого из простых символов — свойство обозначать некоторый класс производных основ. Теперь мы познакомились с другим его свойством — способностью занимать определенное место в комплексе, или, что то же самое, способностью подчинять одни символы и подчиняться другим. Набор этих формальных свойств символа обозначается буквой R , причем R_1 обозначает свойства V , R_2 — свойства N , R_3 — свойства A и R_4 — свойства D .

Эти свойства, называемые реляторами, используются для порождения производных символов от исходных по заданным правилам, которые вместе с правилами аппликации исчерпывают существо аппликативного компонента модели (см. табл. 10).

Таблица 10

	V	N	A	D
R_1	R_1V	R_1N	R_1A	R_1D
R_2	R_2V	R_2N	R_2A	R_2D
R_3	R_3V	R_3N	R_3A	R_3D
R_4	R_4V	R_4N	R_4A	R_4D

R_1V обозначает глагол в позиции глагола; R_1N — существительное в позиции глагола (ср. *был либералом, либеральничал*); R_1A — прилагательное в позиции глагола (ср. *хлеб был свеж и душист*); R_1D — наречие в позиции глагола (в русском языке, по-видимому, неинтерпретируемый случай); R_2V обозначает глагол в позиции существительного (*писатель пописывает, читатель почитывает, чай пить* — не *дрова рубить*); R_2N — существительное в позиции существительного; R_2A — прилагательное в позиции существительного (ср. *величина горы поразила его, багровый и белый отброшен и скомкан*); R_2D — наречие в позиции существительного (тоже, по-видимому, неинтерпретируемый случай в русском языке) и т. д. Следует подчеркнуть, что таким образом моделируются не только механизмы словообразования, но и механизмы словоизменения; не только образование частей речи, но и образование членов предложения. Так, символ R_4N (существительное в позиции наречия) является конструктивным аналогом не только наречий, образованных от существительных (ср. *вдаль, внизу, сверху, домой*), но и существительных в функции обстоятельств (*Петр брал лесом целый час*;

Он говорит без **ошибок**), а также существительных в функции дополнений (*Петр подарил **ребенку игрушку***). Понятия словообразования и словоизменения, обстоятельства и дополнения, принципиально и, надо сказать, вопреки интуиции отождествляемые на идеальном уровне, могут быть различены при интерпретации модели.

Очевидно, что, применив, реляторы R_1 , R_2 , R_3 и R_4 к четырем неприводимым основам N , V , A , D , мы можем на втором шаге этого процесса применить их к шестнадцати основам первой степени производности и получить 64 основы второй степени производности, ср. R_1R_2V (*учит* → *учитель* → **учительствует**), R_2R_3V (*зреет* → *зрелый* → **зрелость**, *переводит* → *переводимый* → **переводимость**), R_4R_3V (*зреет* → *зрелый* → **зрело**, *запоздывает* → *запоздалый* → **запоздало**) и т. д. Можно сформулировать одно рекурсивное правило порождения производных основ: 1) N , V , A , D суть основы; 2) если X — основа, то и R_mX (где $m=1, 2, 3$ или 4) также основа <209>.

До тех пор пока у нас не было правил порождения производных основ и пока каждое правило аппликации могло применяться, по условию, только однажды, модель была ограничена в своих возможностях: она не порождала ни одного действительно интересного комплекса (ср. примеры на стр. 225). Теперь же мы располагаем необходимыми средствами для того, чтобы снять эти ограничения и построить аналог практически любого типа слова, словосочетания и предложения. Эта задача сводится к тому, чтобы указать правила осложнения комплексов.

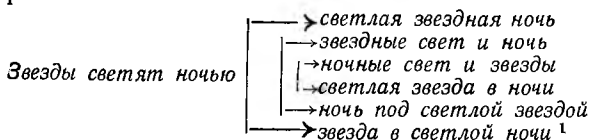
Элементарные комплексы могут быть осложнены несколькими способами. Во-первых, каждый из 14 элементарных комплексов мы можем превратить в неэлементарный, заменяя те или иные неприводимые основы производными, например $R_2XR_1XR_4X$ (ср. *Сторож охраняет магазин*), $R_2XR_1XR_4R_2X$ (ср. *Старуха позеленела от злости*) и т. п. Во-вторых, каждое правило аппликации можно применить не один раз, как это было сделано при порождении 14 элементарных комплексов, а n раз, причем n заранее никак не фиксировано, например $R_3XR_3XR_2X$ (ср. *шерстяной пиджак в клетку*), $R_2XR_1XR_4XR_4XR_4X$ (ср. *Он вменяет мне в вину опоздание; Он внезапно распахнул дверь настежь*) и т. п. В-третьих, неэлементарные комплексы можно получать по следующему рекурсивному определению: 1) элементарные комплексы (14 выражений) суть

комплексы; 2) если X — комплекс, то и $R_m X$ — также комплекс. Из этого определения следует, что релятор может относиться не только к одной основе, но и к ряду основ, связанных отношением подчинения; иными словами, функцию того или иного класса может выполнять не только основа слова, но и «основа» словосочетания и предложения. Таким образом, это правило позволяет нам строить аналоги любых типов так называемых простых распространенных предложений, сложных предложений различных типов и т. п.; ср. комплекс $R_2 X R_1 X R_4 (R_4 X R_3 X R_2 X)$, соответствующий предложению типа *Сторожка стоит в очень глухом лесу*, комплекс $R_2 X R_1 X R_4 (R_2 X R_1 X R_4 X)$, соответствующий предложению *Он сказал, что поезд приходит ночью*, и т. п.

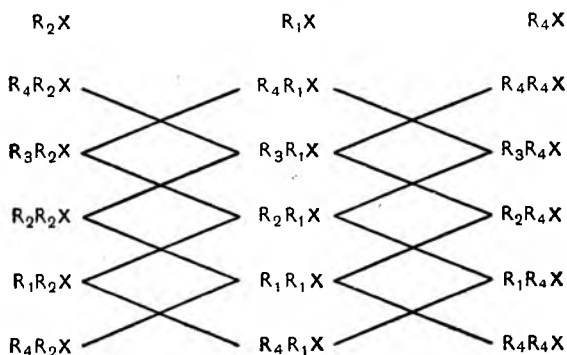
Убедившись, что изложенные выше правила образования обеспечивают порождение, на идеальном уровне, достаточно разнообразных аналогов слов, словосочетаний и предложений, мы перейдем ко второму компоненту аппликативной порождающей модели, который содержит, как сказано выше, трансформационные правила преобразования комплексов. В модели С. К. Шаумяна, в отличие, например, от модели Н. Хомского, эти правила не задаются извне, а вырабатываются на основе имеющейся информации. Средством для выработки этих правил служит специальное исчисление трансформаций.

Трансформации делятся на связанные и несвязанные. Считается, что комплекс A' образован из комплекса A связанной трансформацией, если: 1) для любой основы X_i , входящей в A , в A' найдется основа X_j , образованная из X_i по правилам порождения производных основ (см. таблицу на стр. 227), и наоборот, для каждой основы X_j , входящей в A' , в A найдется основа X_i , от которой образована X_j ; 2) для любой пары основ X_i и Y_i , входящих в A и связанных отношением подчинения, в A' найдется пара соответствующих основ X'_i и Y'_i , также связанных отношением подчинения, и наоборот (ср. определение трансформации на стр. 155). Примеры связанных трансформаций: $R_3 X R_2 X \rightarrow R_2 R_3 X R_3 R_2 X$ (*высокая гора* \rightarrow *высота горы*); $R_2 X R_1 X \rightarrow R_3 R_1 X R_2 X$ (*поезд идет* \rightarrow *идущий поезд*); $R_2 X R_1 X R_4 X \rightarrow R_2 R_1 X R_3 R_4 X R_3 R_2 X$ (*мальчик читает книгу* \rightarrow *чтение книги мальчиком*). Преобразование называется несвязанным, если выполнено только первое из двух сформулированных выше условий.

Пример:



Чтобы показать, каким образом находятся все возможные трансформы данного комплекса, рассмотрим следующий пример. Возьмем трехчленный комплекс $R_2XR_1XR_4X$ (ср. *Писатель завершает книгу*), выпишем в столбик для каждой из входящих в него основ все основы следующей степени производности (для удобства повторив внизу первую строчку) и соединим пары основ в соседних столбиках таким образом, чтобы в результате получались парные комплексы, члены которых могут быть связаны отношением подчинения:



Переходя от каждого из элементов левого столбца к какому-нибудь элементу правого по одному из возможных путей, мы получим все трансформы исходного комплекса $R_2XR_1XR_4X$ первой степени производности. Всего, как легко подсчитать, он имеет 14 трансформов. Разумеется,

¹ Легко заметить, что к числу несвязанных относятся и преобразования типа *Критик организует группу* → *Организация группирует критиков* → *Группа критикует организаторов*. Таким образом, несвязанные преобразования моделируют тот аспект владения грамматикой языка, который проявляется в отмеченной Ч. Фризом способности человека «понимать» фразы типа *Woggles ugged diggles* и строить по их образцу другие подобные фразы (см. стр. 103).

не все они получают интерпретацию в русском языке; к числу интерпретируемых трансформов относятся

$R_2R_2XR_3R_1XR_4R_4X$
 $R_4R_2XR_3R_1XR_2R_4X$
 $R_2R_4XR_1R_1XR_4R_2X$
 $R_3R_2XR_2R_1XR_3R_4X$

Писатель, завершающий книгу
Книга, завершенная писателем
Книга завершается писателем
Завершение книги писателем и т. д.

Ключевыми в этой системе являются трансформации бинарных комплексов, называемые элементарными. Оказывается, что если у нас есть трансформации для каждого из четырех бинарных (двучленных) комплексов, то трансформации многочленных комплексов мы можем определить через эти элементарные трансформации. Действительно, приведенный выше трехчленный трансформ $R_2R_2XR_3R_1XR_4R_4X$ является (теоретико-множественным) объединением двух бинарных трансформов $R_2R_2XR_3R_1X \cup R_3R_1XR_4R_4X$; аналогичным образом

$$R_4R_2XR_3R_1X \cup R_3R_1XR_2R_4X = R_4R_2XR_3R_1XR_2R_4X,$$

$$R_2R_4XR_1R_1X \cup R_1R_1XR_4R_2X = R_2R_4XR_1R_1XR_4R_2X$$

и т. д. Отсюда следует тот важный вывод, что для получения любой трансформации любого комплекса достаточно построить исчисление трансформаций для бинарных комплексов.

Если обозначить надстрочным символом n степень производности некоторой основы и записать произвольный двучленный комплекс как $R_jX R_kX$, то все трансформы этого комплекса можно представить в обобщенном виде следующим образом:

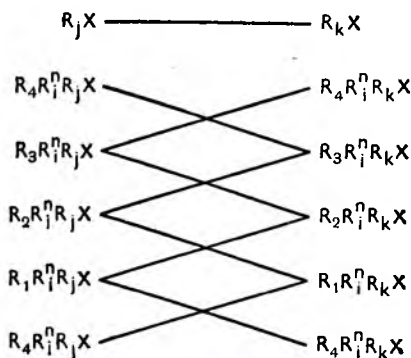


Рис. 15.

Это и дает нам то абстрактное исчисление трансформаций, которое в дальнейшем направляет наши поиски конкретных интерпретируемых трансформаций. Приведем несколько примеров для бинарного комплекса R_3AR_2N :

1. $R_3AR_2N \rightarrow R_4R_3AR_3R_2N$
 - а) *высокая гора* \rightarrow *высокогорный*
широкие плечи \rightarrow *широкоплечий*
 - б) *глубокое несчастье* \rightarrow *глубоко несчастный*
2. $R_3AR_2N \rightarrow R_3R_3AR_4R_2N$
 - а) *широкие плечи* \rightarrow *широкий в плечах*
 - б) *слабое здоровье* \rightarrow *слабый здоровьем*
3. $R_3AR_2N \rightarrow R_2R_3AR_3R_2N$
 - а) *высокая гора* \rightarrow *высота горы*
 - б) *высокая гора* \rightarrow *горная высота*
4. $R_3AR_2N \rightarrow R_2R_3AR_1R_2N$
высокая гора \rightarrow *высота была горою*
5. $R_3AR_2N \rightarrow R_1R_3AR_2R_2N$
высокая гора \rightarrow *гора высится*
6. $R_3AR_2N \rightarrow R_1R_3AR_4R_2N$
широкие плечи \rightarrow *(был) широк в плечах*
7. $R_3AR_2N \rightarrow R_4R_3AR_1R_2N$
хороший обед \rightarrow *хорошо пообедал*

На этом мы заканчиваем изложение формального аппарата модели. Что касается правил интерпретации, то, насколько это допускал жанр нашей книги, они давались по ходу изложения формального аппарата модели, и поэтому мы не будем останавливаться на них особо.

Глава 2

СИНТАКСИЧЕСКИЕ МОДЕЛИ АНАЛИЗА

Переходя к рассмотрению моделей анализа, мы, по крайней мере частично, вступаем в область прикладной лингвистики, которая смежна со структурной лингвистикой, но не тождественна ей. Это — достаточно самостоятельная, широкая и перспективная область машинного перевода, автоматического реферирования текстов и других прикладных задач, плодотворно разрабатываемая не только отдельными исследователями, но и

большими коллективами ученых, см. <64>, <114>, <116>, <117>, <13>, <103>, <120>, <121>, <122>, <125>, <134>, <97>, <74>, <75>, <76>, <189>, <98>, <277>, <279>, <298>, <344>, <73>, <50>. Естественно, что в нашем кратком очерке мы не ставили и не могли поставить перед собой задачи нарисовать сколько-нибудь подробную ее картину, тем более что относящиеся к делу исследования в большинстве случаев насыщены техническими деталями, не представляющими интереса с точки зрения занимающей нас темы. Поскольку, однако, большинство непримитивных моделей анализа основаны на весьма интересных лингвистических предпосылках и являются к тому же действующими моделями речевой деятельности человека, они имеют теоретический интерес и должны быть, хотя бы кратко, представлены в нашей книге¹. В дальнейшем мы даем суммарный очерк наиболее интересных подходов к решению общих лингвистических вопросов, возникающих в моделях такого рода. Две модели, одна из которых принадлежит И. А. Мельчуку, а другая И. Лесерфу, рассматриваются более подробно, чем остальные.

Как мы помним (см. стр. 108), задачу моделирования речевой деятельности человека в самом общем виде можно сформулировать следующим образом. Нам известны грамматика и словарь данного языка. Требуется построить алгоритмы «пользования» грамматикой и словарем, достаточные для того, чтобы машина могла «понять» всякое введенное в нее правильное предложение (в идеале — его смысл, для начала — хотя бы его синтаксическую структуру) и построить правильное предложение по введенной в нее смысловой или синтаксической информации. В данной главе мы будем рассматривать только синтаксические модели указанного здесь типа; мы ограничим себя еще и в том отношении, что оставим в стороне модели синтеза, которые основаны практически на том же круге идей, что и модели анализа, но разработаны гораздо менее детально.

Существуют четыре основных подхода к решению сформулированной выше задачи автоматического анализа

¹ Читателя, интересующегося деталями, мы отсылаем к превосходным обзорам И. А. Мельчука <12>, <114>; см. также работу Р. Равич <147> и недавно вышедшую книгу И. И. Ревзина и В. Ю. Розенцвейга <154>.

текста (перехода от текста к содержащейся в нем синтаксической информации): последовательный анализ, предсказуемый анализ, поиск опорных точек и метод фильтров. Некоторые из них довольно близки друг другу и базируются на идеях, с которыми мы в той или иной мере уже знакомы.

ПОСЛЕДОВАТЕЛЬНЫЙ АНАЛИЗ

Рассмотрим следующее предложение из перевода романа Мельвиля «Моби Дик»: «Акт уплаты представляет собой, я полагаю, наиболее неприятную кару из тех, что навлекли на нас двое обирателей яблоневых садов». Каждый, кто владеет русским языком, поймет по крайней мере прямой смысл этого предложения. Однако его скрытый, аллегорический смысл будет доступен лишь тому, в чьей памяти хранится библейская легенда об Адаме и Еве и древе познания добра и зла, с которого Ева сорвала запретный плод и тем самым навлекла на человечество всяческие бедствия. Тот, кто не знает этой легенды, не поймет более глубокого и в данном контексте единственно важного смысла предложения. Аналогичным образом возникает непонимание и на других уровнях. Очень часто его причиной является то обстоятельство, что мы не знаем слова или (например, при разговоре на иностранном языке) грамматического оборота, употребленного нашим собеседником: в нашей памяти не хранится образца («эталона»), с которым можно было бы сравнить и отождествить поступившее в «устройство понимания» слово (или оборот).

Эти примеры показывают, что в некотором, достаточно узком, но интересном для нас смысле процесс понимания (анализа) текста можно представить как процедуру сопоставления поступающей информации с информацией, хранящейся в памяти, и отождествления совпадающих частей информации.

В сущности, эта идея сопоставления поступающих в устройство понимания единиц с образцами, хранящимися в его памяти, и легла в основу методики последовательного автоматического анализа текста, основные принципы которой были сформулированы в 1954 году В. Ингве <73>, <114>. По мысли В. Ингве, анализирующее устройство, т. е. элек-

тронная вычислительная машина, должно хранить в памяти список (словарь) типичных для данного языка синтагм, записанных в виде последовательностей классов слов. Синтагмы рассматриваются в качестве своего рода эталонов; синтаксическая структура некоторого предложения отыскивается машиной в результате последовательного сопоставления различных цепочек словоформ данного предложения с эталонами, хранящимися в словаре, причем в качестве термина для сравнения каждый раз выбирается максимально длинная синтагма, целиком вкладывающаяся в анализируемое предложение. Найденным словосочетаниям приписываются номера соответствующих синтагм. После обнаружения всех словосочетаний данного предложения устанавливаются отношения между ними (иерархия словосочетаний), и синтаксический анализ предложения считается исчерпанным <114>.

Предложенная В. Ингве методика отличается двумя главными достоинствами: «1. Синтаксический анализ сводится к небольшому числу однообразных, логически очень простых операций, которые осуществляют отыскание в тексте последовательностей из словаря. 2. Правила синтаксического анализа — одни и те же для любого языка. От языка к языку меняются лишь перечни типовых словосочетаний, а правила отыскания словосочетаний с помощью этих перечней остаются одинаковыми» <114, 9—10>.

Наиболее последовательно обрисованный здесь подход к решению проблемы автоматического анализа текста проведен в алгоритме И. А. Мельчука, выходящем далеко за рамки тех прикладных задач машинного перевода, для которых он был первоначально предназначен. Рассмотрению этого алгоритма мы и посвятим данный раздел¹.

Для понимания модели И. А. Мельчука (и других моделей того же типа) существенно уяснить различие между объектами, которые она обрабатывает, и объектами, с помощью которых она производит эту обработку. Первые

¹ Мы отнесли модель И. А. Мельчука к числу алгоритмов последовательного анализа, так как процесс анализа каждой фразы разворачивается в ней в основном слева направо. Однако переход от реального текста к изображению его синтаксической структуры не обязательно завершается за один прогон. Фраза может просматриваться машиной слева направо несколько раз, причем каждый раз устраняются ошибки, которые могли быть допущены на предыдущих этапах анализа. Этим модель И. А. Мельчука отличается от алгоритмов последовательного анализа в строгом смысле слова.

называются объектами анализа, а вторые — единицами или элементами анализа. Объекты анализа суть те или иные единицы текста, вводимого в машину, т. е. речевые произведения; к их числу относятся словоформы, сегменты и фразы (определения см. ниже). Элементы анализа суть те единицы языка, которые хранятся в памяти машины и используются ею как средство анализа текста. К числу важнейших единиц языка относятся морфы (основы и аффиксы) и синтагмы (классы двучленных сочетаний словоформ или сегментов, имеющих одинаковое грамматическое строение). Анализ, как и у В. Ингве, состоит в сопоставлении, по определенным правилам, вводимых в машину текстовых объектов с хранящимися в ее памяти языковыми объектами и приписывании текстовым объектам, в зависимости от результатов сопоставления, той или иной информации. Пользуясь несколько грубой, но наглядной аналогией, мы могли бы уподобить хранящуюся в памяти машины информацию о единицах языка грамматическим знаниям человека, а правила анализа — тем автоматическим навыкам использования грамматических знаний, которые позволяют человеку понимать текст.

Процедура анализа текста распадается на две большие части — предварительное препарирование текста, осуществляемое вспомогательными алгоритмами, и собственно анализ, проводимый основным алгоритмом. Входом для вспомогательных алгоритмов является реальный письменный текст, а выходом — так называемые гипотетические (исходные) объекты текста, к числу которых относятся гипотетические словоформы (знаки препинания или последовательности букв, ограниченные пробелами), гипотетические сегменты (последовательности словоформ, ограниченные знаками препинания или союзами) и гипотетические фразы (последовательности сегментов, ограниченные точечными знаками). Входом для основного алгоритма является препарированный текст, разбитый на гипотетические словоформы, сегменты и фразы, а выходом — синтаксическое дерево зависимостей или какой-либо другой подобный ему формальный объект, в явной форме изображающий синтаксическую структуру каждого из предложений. На том или ином этапе в процессе перехода от первоначально препарированного предложения к его

синтаксическому дереву каждый гипотетический объект превращается в окончательный, или приведенный. Окончательная словоформа — это аналог слова или фразеологической единицы; окончательный сегмент — это аналог простого предложения (т. е. простое предложение, деепричастный, причастный и другие обороты); окончательная фраза соответствует полному предложению.

Рассмотрим в качестве примера, иллюстрирующего эти понятия, результаты обработки предложения *Так как подобный шток не годится ни для какого цилиндра, схема, изображенная на рис. 4, должна быть заменена другой.* Словоформы в обработанных предложениях отделены друг от друга пробелами; границы между сегментами показаны либо вертикальными чертами, либо римскими цифрами; границы фразы соответствуют две вертикальные черты.

Г и п о т е т и ч е с к и е о б ъ е к т ы :

. Так как подобный шток не годится ни для какого цилиндра|, схема|, изображенная на рис|| . 4| , должна быть заменена другой||

О к о н ч а т е л ь н ы е (п р и в е д е н н ы е)
о б ъ е к т ы :

I таккак подобный шток негодится для никакого цилиндра

II схема должна быть заменена другой

III изображенная на рисунке 4 ||

Основной алгоритм представляет собой множество таблиц стандартной формы, в которых собраны все сведения о том или ином языке, и правил обращения с таблицами. Таблицы — это своеобразные словари, где собраны элементарные единицы рассматриваемого языка: морфы для морфологического анализа и синтагмы для синтаксического. Правила обращения с таблицами образуют алгоритм в собственном смысле слова. Они одинаковы для всех привлекаемых языков (при условии, что конкретные данные этих языков собраны в таблицах стандартной формы) и зависят исключительно от формы таблиц. В этом отношении их можно сравнить с инструкцией по обращению с алфавитными словарями, не изменяющейся от языка к языку и зависящей только от порядка букв в алфавите <118>.

Алгоритм делится на два больших блока: блок морфологического анализа и блок синтаксического анализа.

Объектом морфологического анализа являются (исходные) словоформы. В процессе анализа каждая словоформа изменяемого слова сопоставляется с морфами, хранящимися в словаре морф, и, в случае если находится последовательность морф, совпадающих с какой-то частью словоформы (вкладываемых в нее), последняя разрезается на морфы. При этом учитывается возможность омонимии и другие трудные для анализа явления.

«Словарная статья» морфологического словаря, с помощью которого осуществляется поиск морф и разрезание словоформ, представляет собой строчку, разделенную на пять частей. В части I помещается порядковый номер морфы, в части II — сама морфа, а в частях III и IV — информация к ней, представляющая в явной и удобной форме все ее грамматически существенные свойства (например, информация о времени, числе, лице и т. п. для глагольных окончаний). Часть V отведена для «переадресации» — указания дальнейших действий, которые должны выполняться после отсечения данной морфы.

Важнейший этап морфологического анализа состоит в замене морф данной словоформы цепочками информации к ней, которые в совокупности дают характеристику словоформы, обеспечивающую дальнейший анализ. Цепочки информации к словоформам и являются выходом морфологического блока алгоритма.

Информации к словоформам вместе с информацией об исходных (гипотетических) сегментах и фразах подаются на вход синтаксического блока алгоритма, который на этой основе обнаруживает (конструирует) синтаксическую структуру предложения. Синтаксический анализ распадается на два этапа: на первом этапе (внутри-сегментный анализ) устанавливаются синтаксические связи словоформ внутри одного сегмента; на втором этапе (межсегментный анализ) отыскиваются синтаксические связи между сегментами.

Для того чтобы читателю легче было следить за нашим изложением, мы сначала скажем несколько слов о способе изображения результатов синтаксического анализа (т. е. форме, в которой ищется ответ на поставленный выше вопрос), а затем опишем процесс обнаружения синтаксической структуры предложения.

Синтаксическая структура предложения изображается

деревом зависимостей с нумерованными стрелками, каждая из которых обозначает некоторое отношение непосредственной доминации. С общим понятием бинарного отношения доминации (подчинения) мы уже знакомы; однако в модели И. А. Мельчука рассматривается не одно отношение доминации и даже не 2—5, как это принято в других системах, а 31, т. е. на порядок больше. Это достигается расщеплением некоторых подчинительных связей на ряд более дробных и тонких связей. Так, в алгоритме И. А. Мельчука рассматривается не одно объектное отношение, а 3. Первое объектное отношение иллюстрируется

примерами: *решать задачу*, *хранение информации*, *принадлежащий к множеству*, *избегает ошибок*; второе — примерами: *приписать букве индекс*, *сведение вычисления к сложениям*; третье — примером: *сравнить самолет с ракетой по скорости*. Вместо одного определительного отношения рассматривается около десятка различных: определительное в собственном смысле слова (*обычная запись*, *вычислительный процесс*, *действовать машинально*, *чисто автоматически*); указательное (*этот человек*); притяжательное (*орудие вычислителя*); кванторное (*все данные*, *любая буква*, *некоторые процессы*); общекваликативное (*такие таблицы*); порядковое (*первое слово*); количественное (*пять страниц*); субстантивно-аттрибутивное (*отверстие диаметром*); партитивное (*элемент множества*); общегенетивное (*лист бумаги*, *пример алгоритма*) и т. д.

Столь детальная классификация отношений непосредственной доминации (опознаваемых алгоритмом автоматически) позволяет изобразить не только синтаксическую структуру предложения, но и значительную часть его смысла.

Средством, с помощью которого осуществляется синтаксический анализ текста, является словарь конфигураций. **К о н ф и г у р а ц и я** — это синтагма или подобная ей единица языка, записанная в стандартной форме. Как и словарная статья морфологического словаря, конфигурация представляет собой строку, разделенную на пять частей. В части I помещается ее порядковый номер, в части II — синтагма или другая языковая единица синтаксического уровня, а в частях III и IV — информация о синтагме, в частности информация о типе (номере) подчинительного отношения, связывающего элементы синтагмы. Часть V предназначена для «переадресации»; в нее записываются номера конфигураций, к которым следует обратиться после данной.

Синтагма — основная единица анализа — представляет собой, как помнит читатель, класс двучленных сочетаний словоформ или сегментов, имеющих одинаковое грамматическое строение. Грамматическое строение синтагмы описывается в терминах некоторых, преимущественно морфологических, категорий; ср. синтагму «переходный глагол+существительное в винительном падеже — в любом порядке» (*рассматривает цели*) или синтагму «отглагольное существительное от переходного глагола+существительное в родительном падеже — в прямом порядке» (*рассмотрение целей*) и т. п. Поскольку синтагма определяется в терминах преимущественно морфологических категорий, а приписанное ей отношение непосредственной доминанции — в терминах синтаксических или семантических категорий, между синтагмами и типами подчинительных отношений нет взаимнооднозначного соответствия. Разным синтагмам может соответствовать одно и то же отношение (ср. примеры первого объектного отношения на стр. 239), а разным отношениям может соответствовать одна и та же синтагма (ср. *принадлежать к множеству*, где выражается первое объектное отношение, и *свести к сложениям*, где выражается второе объектное отношение).

Объектом синтаксического анализа на первом этапе (внутрисегментный анализ) являются исходные сегменты. Существенно, что каждая словоформа сегмента представлена цепочкой информации того же самого типа, который используется для описания грамматического строения синтагм. В процессе анализа каждый сегмент сопоставляется с синтагмами и другими языковыми единицами синтакси-

ческого уровня. Если какая-то часть сегмента совпадает с определенной синтагмой, то она отождествляется с этой синтагмой и на нее переносится приписанная синтагме информация о типе подчинительного отношения. При этом вырабатываются сведения о сегменте в целом (а иногда даже о фразе). После отождествления всех синтагм данного сегмента мы получаем некоторый куст строящегося дерева зависимостей, в котором главный и подчиненный элементы каждой синтагмы связаны нумерованной стрелкой.

После того как для всех сегментов закончен внутрисегментный анализ, выработанная информация подается на вход второго — межсегментного — этапа синтаксического анализа. На этом этапе аналогичным образом (т. е. путем сопоставления со словарем синтагм) устанавливаются связи между сегментами и устраняются ошибки внутрисегментного анализа, которые могли быть ранее допущены. Результатом межсегментного анализа является синтаксическое дерево предложения с нумерованными стрелками, обладающее следующими формальными свойствами: 1) в нем есть ровно одна главная вершина; 2) оно связно в том смысле, что из главной вершины можно пройти по стрелкам зависимостей в любой узел дерева; 3) у каждого подчиненного элемента, не являющегося сегментом, может быть только один «хозяин» (что касается сегмента, например, простого предложения, то он может одновременно зависеть и от двух элементов — другого сегмента и одной из его словоформ). Из этого, между прочим, следует, что в дереве, изображающем синтаксическую структуру сегмента, не может быть циклов.

Как было сказано выше (стр. 235), процесс анализа текста в модели И. А. Мельчука и, в особенности, в алгоритмах последовательного анализа в строгом смысле слова, разворачивается слева направо и строится на основе того материала, который реально представлен в данной фразе; в ходе анализа практически не выдвигается никаких гипотез о том, что во фразе могло бы быть. Такой способ анализа в общем соответствует представлению, что человек понимает текст по мере его поступления в «анализирующее устройство» мозга; при этом формы и слова, которые фактически появляются во фразе, ожидаются человеком не больше, чем какие-нибудь другие формы и слова.

В излагаемых ниже алгоритмах формализуется несколько иное содержательное представление о том, каким

образом протекает процесс понимания текста человеком. Предполагается, что, извлекая информацию из текущей формы, человек строит некоторое число гипотез о том, какие другие формы (и смыслы) могут встретиться в той же фразе; он о ж и д а е т появления одних форм (слов, смыслов) и не ожидает появления других, причем мера понимания текста определяется тем, в какой степени оправдываются его ожидания.

ПРЕДСКАЗУЕМОСТНЫЙ АНАЛИЗ

Чтобы пояснить основную идею предсказуемого автоматического анализа, главные принципы которого были содержательно изложены Ч. Хоккетом несколько лет назад <311>, а затем формализованы И. Родес <114> и А. Эттингером <344>, <114>, проведем следующий эксперимент: попытаемся догадаться, какое слово может заполнить пробел во фразе *Весь декабрь было тепло, и шли дожди; лишь в январе выпал первый* —. Решение, по-видимому, однозначно: на последнем месте в этой фразе, вероятнее всего, будет стоять слово *снег*. Появление другого слова вызвало бы у нас удивление, а появление слова *снег* лишь подтверждает ту гипотезу, которую мы бессознательно сформировали, прочитав уже «поступившие» слова. Такие гипотезы мы постоянно строим и для других уровней, в частности грамматического. Если в начале предложения стоит артикль или предлог, мы ожидаем появления существительного; если появилось существительное в именительном падеже, мы ожидаем появления глагола в личной форме; если появился переходный глагол, мы ожидаем появления существительного в винительном падеже и т. д. Важным для нас является то обстоятельство, что уже поступившая часть предложения ограничивает возможности варьирования грамматической структуры в остальной части; именно поэтому мы и можем делать обоснованные предсказания о грамматической структуре непроизведенной части предложения.

Изложенное здесь представление о процессе понимания (анализа) текста формализуется техникой предсказуемого анализа. Средством анализа синтаксической структуры предложения является хранящийся в памяти машины набор синтаксических предсказаний (НСП), содержащий в себе гипотетический

перечень возможных в некотором предложении синтаксических структур. «До начала анализа НСП содержит, например, указания о возможности группы подлежащего, о возможности сказуемого, о возможности группы дополнения, о возможности различных обстоятельств и некоторые вспомогательные указания. При каждом таком «предсказании» указываются свойства (морфологические и лексические) тех слов, которые могут удовлетворять этому предсказанию, т. е. выполнять соответствующую синтаксическую функцию. В ходе анализа НСП изменяется: одни предсказания «сбываются» и устраняются из НСП, другие предсказания уточняются и конкретизируются, а кроме того, в НСП появляются новые предсказания. Например, когда найдено слово, являющееся подлежащим, устраняется предсказание о подлежащем и уточняется предсказание о сказуемом: точно указывается, в каком лице, числе (и роде — в прошедшем времени) должен быть глагол-сказуемое. Одновременно появляется новое предсказание — о возможности приименного дополнения в родительном падеже и т. д.

Процесс анализа заключается в последовательном сравнении информации к словам анализируемого предложения со всеми предсказаниями, перечисленными в НСП. Когда для слова найдено «подходящее» предсказание, слову приписывается пометка соответствующей синтаксической функции, а в НСП вносятся необходимые изменения» <114, 14>. Анализ продолжается до тех пор, пока каждой словоформе предложения не будет приписана определенная синтаксическая функция.

ПОИСК ОПОРНЫХ ТОЧЕК

Этот метод представляет собой дальнейшее развитие идей, лежащих в основе предсказуемого анализа <277>, <114>, и близок к «анализу по цепочкам» (String Analysis), разрабатываемому З. Харрисом <296> и Р. Лонгакром <332>.

Поиск синтаксической структуры предложения начинается с обнаружения его **о п о р н ы х т о ч е к**. Опорной точкой сложного предложения считается главное предложение; опорной точкой простого предложения — его сказуемое. Именно эти точки содержат максимальное количество грамматической информации; их обнаружение

позволяет нам выдвигать обоснованные гипотезы (предсказания), касающиеся других элементов предложения. Так, в русском языке число, лицо и — в прошедшем времени — род сказуемого образуют в совокупности достаточную информацию, чтобы найти подлежащее предложения, с которым, как известно, сказуемое согласуется по этим признакам. Кроме того, глагольное сказуемое характеризуется обычно определенным типом управления, по которому можно найти его дополнения. Так, глагол *принадлежать* надежно предсказывает дополнение в дательном падеже (иногда с предлогом *к*), глагол *стремиться* — дополнение в дательном падеже с предлогом *к* или инфинитив, глагол *интересовать* — дополнение в винительном падеже и т. д. (см. примеры на стр. 137). Опорной точкой группы существительного является существительное, форма которого (род, число, падеж) предсказывает форму согласованного с ним определения, и т. д.

Интересно отметить, что на тех же принципах построен грамматический разбор предложения на уроках латинского языка: ключом к латинскому предложению испокон веков считается его сказуемое.

Для того чтобы подобный направленный поиск синтаксической структуры предложения мог быть осуществлен, в память машины должен быть введен автоматический словарь, в котором каждому слову приписан некоторый грамматический код, представляющий морфологические свойства слова и указывающий все его потенциальные синтаксические возможности (управление, согласование). Тогда синтаксический поиск сводится к установлению того, какие из присущих слову возможностей фактически реализуются в данном предложении.

Как ясно из сказанного, предсказуемый анализ и метод опорных точек имеют то общее свойство, что обнаружение синтаксической структуры предложения является в них результатом последовательной проверки ряда локальных гипотез: на каждом шаге анализа выдвигается **о г р а н и ч е н н о е** число гипотез и, если одна из них оправдывается, алгоритм вырабатывает соответствующий ей фрагмент информации о синтаксической структуре предложения. Рассматриваемый ниже метод фильтров сближается с двумя только что описанными методами в том отношении, что в нем поиск синтаксической структуры предложения начинается с выдвижения и проверки

некоторого множества гипотез; однако его принципиальное отличие от предсказуемого анализа и метода опорных точек состоит в том, что в начале работы множество проверяемых гипотез никак не ограничено.

МЕТОД ФИЛЬТРОВ

Основные принципы метода фильтров были сформулированы в работах И. Лесерфа <98> и развиты рядом других исследователей, среди которых особого упоминания заслуживает Л. Н. Иорданская <75>, <76>. Эти принципы можно суммировать следующим образом: задача заключается в том, чтобы найти синтаксическую структуру текста, т. е. сопоставить каждому содержащемуся в нем предложению некоторое дерево зависимостей. Обнаружение дерева зависимостей начинается с рассмотрения всего множества допустимых решений, каждое из которых является гипотезой о возможных синтаксических функциях словоформ предложения. Затем гипотезы проверяются с помощью специальной программы, и те из них, которые себя не оправдывают, отбрасываются. Программа служит, таким образом, в качестве своего рода фильтра, задерживающего все неправильные решения. На выход пропускается наилучшая в некотором смысле гипотеза, которая и считается решением (ср. стр. 133). Число наилучших гипотез может быть больше одного.

Возьмем в качестве примера множество гипотез, которые можно выдвинуть при анализе фразы *Le pilote ferme la porte* — «Пилот закрывает дверь» <98, 43>. В машинной памяти, хранящей грамматику и словарь данного языка, словоформам этой фразы сопоставлены информации, показанные на схеме 4 (см. стр. 246).

«Для того чтобы представить эту фразу в виде «дерева зависимостей», необходимо проверить каждую из гипотез о грамматической функции каждого отдельного слова. А это подразумевает в свою очередь выбор между $2 \times 3 \times 4 \times 3 \times 3 = 216$ грамматическими гипотезами» <98, 43>, которые и образуют в совокупности множество допустимых решений.

Среди допустимых решений для данной фразы есть два верных: 1. «Пилот закрывает дверь»; 2. «Сильный пилот несет ее». Чтобы отсеять неправильные решения, машина

пускает в ход грамматические фильтры, роль которых могут выполнять, например, правила сочетаемости словоформ в составе предложения, характерные для данного языка. В частности, достаточно одного правила о невозможности последовательности «артикль+глагол в личной форме», чтобы отсеять 56 гипотез из 216.

Схема 4

<i>Le</i>	<i>pilote</i>	<i>jerme</i>	<i>la</i>	<i>porte</i>
1. Артикль 2. Местоимение	1. Существительное 2. Прилагательное 3. Глагол	1. Существительное 2. Прилагательное 3. Глагол 4. Наречие	1. Артикль 2. Местоимение 3. Существительное	1. Существительное 2. Прилагательное 3. Глагол

Кроме этих частных ограничений, характеризующих грамматический строй данного языка, для устранения неправильных гипотез используется еще один, гораздо более сложный и интересный, фильтр, построенный с учетом универсальных синтаксических закономерностей, характеризующих правильную синтаксическую структуру в очень большом числе языков. Эти закономерности, систематическое исследование которых началось лишь в самое последнее время <98>, <298>, <189>, <50>, <75>, <76>, <133>, представляют большой лингвистический интерес и будут кратко описаны ниже.

Рассмотрим следующие фразы:

1) Допущенные нарушения Кобловым могли привести к крушению пассажирского поезда («Известия», 12/VIII—64).

2) Нарушившая сторона настоящий договор... обязана уплатить другой стороне штраф (Из персонального обязательства, «Правда», 5/VII — 64).

3) Нарушившие настоящий договор вносят деньги сразу же после нарушения, которые подлежат израсходованию... (Т а м ж е).

4) Я ножницами боюсь когда они играют <216>.

5) Это были преступники, отрезанные ломти от общества (Ф. М. Достоевский).

При взгляде на эти фразы нас останавливает их очевидная неграмматичность, связанная с какими-то нарушениями привычного порядка слов. Следовало бы сказать *Допущенные Кобловым нарушения...*; *Нарушившая настоящий договор сторона ...*; *...вносят сразу же после нарушения деньги, которые подлежат израсходованию*; *Я боюсь,*

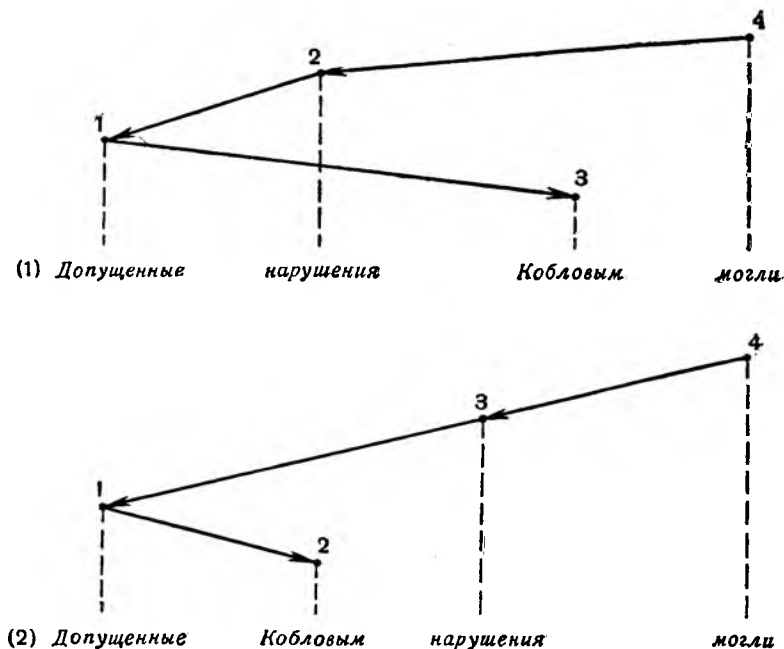


Рис. 16.

когда они играют ножницами; Это были преступники, отрезанные от общества ломти. Заметим, что существо наших поправок сводится к тому, что мы помещаем зависимое слово ближе к главному и добиваемся того, что между главным и зависимым не остается никаких других слов, которые не были бы связаны, непосредственно или опосредствованно, с одним из них.

Посмотрим, нельзя ли описать точно природу того типа неграмматичности, который возникает при нарушениях

нормального порядка слов, подобных только что рассмотренным. Возьмем какое-либо предложение в двух его вариантах — правильном и неправильном, нарисуем для каждого из них синтаксическое дерево и опустим из каждой точки дерева перпендикуляр на воображаемую горизонтальную линию (рис. 16).

На рис. 16 (1) перпендикуляр, опущенный из точки 2, пересекает ветвь 1—3. На рис. 16 (2) таких пересечений нет. Дерево, в котором никакая ветвь, отходящая от некоторой точки *i*, не пересекает перпендикуляра, опущенного из более высокой точки, называется *проективным* (или *конфигурационным* <189>). В силу этого определения дерево (2) проективно, а дерево (1) непроективно. Оказывается, что неграмматичность всех приведенных выше предложений коренится в непроективности их деревьев; правильные синтаксические структуры должны быть проективны.

Поскольку свойство проективности было открыто сравнительно недавно, наука пока не накопила достоверных данных, характеризующих языки с этой точки зрения; немногие имеющиеся сведения как будто подтверждают предположение о безусловном преобладании проективных структур в большинстве языков. Так, И. Лесерфу удалось установить, что французский язык является проективным практически на 100%. То же самое, по-видимому, справедливо для немецкого, голландского, итальянского и ряда других языков <98>. В специальных исследованиях, посвященных русскому языку, на 30 000 фраз научного текста было найдено только две непроективные. Если в некоторых языках, например латыни и русском, обнаруживаются отклонения от этого правила, то их всегда можно перечислить. В частности, для русского языка уже сделаны пробные перечни типов допустимых непроективных структур <75>, <76>, <133>¹.

¹ Эти структуры можно иллюстрировать следующими примерами <133>: 1) *лучший поэт в мире* (ср. проективную структуру *лучший в мире поэт*), 2) *более ценный, чем золото* (ср. *ценный более, чем золото*), 3) *должен был сказать* (ср. *был должен сказать*), 4) *его нельзя понять* (ср. *нельзя понять его*), 5) *примером может служить* (ср. *может служить примером*) и ряд других. В пословицах и иных поэтических произведениях непроективность встречается чаще, ср. *Муж жену любит здоровую, а брат сестру — богатую*. Интересные примеры допустимой непроективности можно найти у А. М. Пешковского <144, 296—297>.

Если непроективные синтаксические структуры, за некоторыми разрешенными и точно определяемыми исключениями, действительно являются неправильными, то требование проективности можно использовать в качестве исключительно мощного фильтра, устраняющего из множества допустимых решений большую часть неправильных. Для этого достаточно ввести в машину специальный блок, пропускающий для дальнейшего анализа только проективные структуры. Интересно отметить, что «коэффициент полезного действия» фильтра возрастает по мере увеличения числа слов в предложении <98>, <298>. И. Лесерф приводит следующие простые расчеты, иллюстрирующие это утверждение (табл. 11).

Таблица 11

Число вершин в дереве	Число проективных деревьев	Число непроективных деревьев	Общее число деревьев	Процент непроективных деревьев
1	1	0	1	0
2	2	0	2	0
3	7	2	9	22
4	30	34	64	53
5	143	482	625	77
6	728	7048	7776	90,6
7	3876	113773	117649	96,7

Таким образом, работа первого фильтра, пропускающего для последующего анализа только проективные структуры, устраняет громадное большинство неправильных гипотез. В дальнейшем поиск наилучшего решения из числа тех, которые удовлетворяют требованию проективности, протекает следующим образом. Известно, что проективную структуру можно изображать с помощью скобок. Что касается непроективной синтаксической структуры, то она может быть изображена только в виде дерева, для которого не существует эквивалентной ему скобочной записи. При переводе предложения в скобочную запись используется второй фильтр — перечень синтаксических свойств слова и ограничений на сочетаемость словоформ в данном языке (ср. стр. 246). Учет этих ограничений позволяет найти правильную расстановку скобок и приписать каждому предложению дерево зависимостей (или

дерево составляющих), изображающее его синтаксическую структуру.

При изложенном подходе используется только одно универсальное свойство правильной синтаксической структуры — свойство проективности. Хотя фильтр, основанный на использовании проективности, задерживает большую часть неправильных гипотез, он все же пропускает для дальнейшего анализа несколько тысяч, а при достаточно длинном предложении — несколько десятков и даже сотен тысяч гипотез. Выбор нужной гипотезы в этом множестве осуществляется по правилам, которые должны составляться о с о б о для каждого языка. Естественно возникает мысль о том, нельзя ли построить еще несколько у н и в е р с а л ь н ы х ф и л ь т р о в, которые могут быть использованы для обнаружения правильной синтаксической структуры на л ю б о м языке, с тем чтобы отодвинуть дальше к концу или, во всяком случае, ограничить применение специфических для данного языка фильтров. Совершенно ясно, что такая универсализация системы фильтров, если она возможна, представляет не только практический, но и теоретический интерес: всякая общая теория не только выгоднее (экономичнее) частной, но и более верна по существу, так как ближе подходит к реальному единству своего объекта.

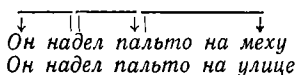
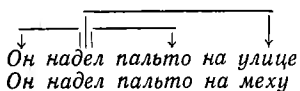
Наиболее существенные результаты на этом пути были получены в работах Л. Н. Иорданской <75>, <76>, проанализировавшей, помимо проективности, некоторые другие общие свойства правильных структур и предложившей алгоритм автоматического синтаксического анализа, в котором используется построенная на учете этих свойств более общая система фильтров.

Правильная синтаксическая структура обладает, по определению, следующими свойствами. Во-первых, она должна быть представима в виде дерева зависимостей, обладающего описанными на стр. 241 свойствами, или должна удовлетворять каким-то другим строго сформулированным соглашениям относительно способа представления синтаксических структур. Во-вторых, она должна удовлетворять ограничениям на сочетаемость синтаксических связей; одно из них — требование проективности дерева. В-третьих, в ней должно быть определенное соответствие между парами связанных элементов и знаками препинания. В-четвертых, в ней должны быть соблю-

дены характерные для данного языка ограничения, налагаемые на сочетаемость элементов. В русском языке, например, предлог в правильной структуре не должен быть связан с инфинитивом, в то время как во французском и шведском, где этот запрет не действует, установление такой связи не нарушает правильности структуры. Наконец, в-пятых, правильная структура должна быть насыщенной, т. е. в ней, за исключением особо оговариваемых случаев, должны быть представлены все те зависимые элементы, наличие которых предсказывается соответствующим главным элементом. Так, если в правильной структуре есть глагол в личной форме, в ней должно быть и имя в именительном падеже; если в ней имеется сильно управляющий глагол, то должна быть и предсказываемая им форма и т. д.

Синтаксическая структура данной фразы отыскивается следующим образом. Для данной фразы строятся все возможные синтаксические структуры (элементы фразы попарно связываются всеми возможными способами). Затем каждая построенная структура пропускается через фильтры — свойства правильной синтаксической структуры; в результате все неправильные структуры отсеиваются, а правильные структуры подаются на выход.

Интересно, что в общем случае предложению может быть приписано более одного дерева зависимостей, обладающего всеми свойствами правильной структуры. Так, предложение *Поверили его надзору подъячего* адекватно описывается деревом, в котором *его* зависит от *надзору*, а *надзору* — от *поверили*, и деревом, в котором обе эти словоформы зависят от глагола *поверили*. Легко заметить, что это соответствует факту объективной двусмысленности анализируемого предложения. Однако по два взаимоисключающих анализа будет получено и для вполне однозначных предложений типа *Он надел пальто на улице* и *Он надел пальто на меху*:



Между тем совершенно очевидно, что первое дерево описывает структуру первого, но не второго предложения, а второе дерево — структуру второго предложения, но не пер-

вого. Чтобы устранить одно из двух возможных синтаксических решений, надо привлечь для анализа предложения соображения, более сильные, чем чисто синтаксические; иными словами, необходимо обратиться к семантике.

Более сильная, семантическая, модель рассматривается нами в следующей главе.

Глава 3

СЕМАНТИЧЕСКАЯ МОДЕЛЬ АНАЛИЗА

До сих пор мы имели дело с алгоритмами и исчислениями, в которых моделировалась способность человека понимать и переводить «грамматически правильные» предложения, т. е. владение грамматикой языка. Здесь мы рассмотрим модель, имитирующую способность человека понимать и производить «семантически правильные», т. е. осмысленные, предложения. В данном случае моделируется не только владение грамматикой языка, но и владение значением слов. Задача моделирования осмысленного речевого поведения человека была поставлена в наиболее интересной форме в работах А. К. Жолковского, Н. Н. Леонтьевой, Ю. С. Мартемьянова, В. Ю. Розенцвейга, Ю. К. Щеглова и некоторых других сотрудников той же группы <59>, <57>, <110a>¹ (хотя в принципе она ставилась и раньше: <108>, <109>, <6>); на эти работы, а также на материалы рецензии, написанной автором совместно с К. И. Бабицким, мы и будем опираться в дальнейшем изложении.

Попытаемся выяснить в самых общих чертах, в чем проявляется «владение значением слов». Допустим, например, что в разговоре с иностранцем, плохо знающим русский язык, мы сказали *У меня нет уверенности на этот счет*, и это предложение было ему непонятно. Если мы хорошо владеем значениями русских слов, мы сумеем *перевести* это предложение в другие предложения русского языка, т. е. пересказать ту же мысль другими способами, например: *Я не испытываю уверенности на этот счет; Это не внушает мне уверенности; Это не кажется мне*

¹ Значительный интерес представляет также работа Дж. Катца и П. Постала <325a>.

бесспорным; Я не уверен в этом; Вряд ли это так; Это, наверное, не так; Я испытываю сомнения на этот счет; Это вызывает у меня сомнения; Это внушает мне сомнения; Это кажется мне сомнительным; Я сомневаюсь в этом и т. д. По всей видимости, какое-нибудь из приведенных здесь предложений будет понятно нашему собеседнику, и мы, таким образом, сообщим ему ту мысль, которая и была предметом нашего высказывания, хотя форма ее выражения будет отличаться от первоначальной.

С другой стороны, если мы хорошо владем значениями русских слов, мы сумеем установить семантическое тождество всех 12 фраз, т. е. поймем, что во всех этих случаях мы имеем дело с превращениями одной и той же мысли, которая в своей основе остается неизменной при любых способах выражения.

Следовательно, владение значением слов проявляется у говорящего в способности по-разному выразить одну и ту же мысль, а у слушающего — в способности понять семантическое тождество внешне различных высказываний <110а>.

Говоря о разных способах выражения одной и той же мысли, или о семантическом тождестве внешне различных высказываний, мы имеем в виду, что существует некий не данный нам в прямом наблюдении семантический язык, или «язык мысли»¹. Психологи относительно давно экспериментально установили, что запоминание мысли не требует запоминания слов, которыми она выражена. Мы говорим: «Я не помню его точных слов, но мысль его состояла в следующем», и мы излагаем далее своими словами то, что сохранили в своем сознании на языке мысли. Если допустить существование такого языка, то производство осмысленного предложения можно представить как перевод с семантического языка на естественный, а понимание предложения — как перевод с естественного языка на семантический. Возможность по-разному выразить одну и ту же мысль означает, что у некоторого выражения семантического языка есть несколько переводов на естественный язык.

Разумно предположить, что на семантическом языке мысль имеет единственный стандартный способ записи, и

¹ Ср. идеи Ш. Балли об идеальном интеллектуальном языке-идентификаторе <242>.

если на естественном языке она выражается несколькими различными способами, которые мы признаем равносильными данной записи, это в общем случае значит, что на разные слова приходятся разные части выражаемой мысли. Фразы *Он недомогает* и *Он плохо себя чувствует* выражают одну и ту же мысль, причем та ее часть, которая в первой фразе выражена одним словом, распределена во второй фразе между тремя различными словами. Из анализа этого примера мы можем сделать первый важный вывод: значение слова в общем случае не является элементарной семантической единицей. Оно делимо на более элементарные смыслы, которые, по предположению, и являются единицами словаря семантического языка¹.

Та или иная комбинация элементарных смыслов образует значение некоторого слова естественного языка. Ясно, что небольшое число элементарных смыслов дает очень большое число возможных комбинаций, реализуемых словами естественного языка. Если, например, у нас есть элементарные смыслы «сам», «кто-то», «иметь», «каузировать» («заставлять»), «переставать», «начинать» и «не», то с их помощью мы можем определить довольно большую группу слов русского языка, в которую войдут прежде всего семь выписанных выше слов, поскольку, будучи единицами семантического языка, они одновременно являются и словами русского языка. Кроме того, определения получают такие русские слова, как: 1) *владеть* = «иметь»; 2) *обладать* = «иметь»; 3) *брать* = «заставлять себя иметь»; 4) *давать* = «заставлять кого-то иметь»; 5) *приобретать* = «начинать иметь»; 6) *утрачивать* = «переставать иметь»; 7) *сохранять* = «не переставать иметь»; 8) *лишать* = «заставлять кого-то переставать иметь» и т. п. Как и в случае с трансформационной грамматикой, пониманию существа дела поможет физический (или, может быть, химический) образ: единицы семантического языка — это те элементы («атомы»), из различных комбинаций которых складываются «молекулы» — значения реальных слов естественного языка.

¹ Предположение о разложимости значения слова на элементарные смыслы (дифференциальные семантические признаки) подтверждается некоторыми экспериментальными данными. Весьма интересные в этом отношении материалы опубликованы Вяч. Вс. Ивановым в его исследовании о лингвистической афазии <71>. См. также <6>.

Мы установили, таким образом, что у семантического языка есть некоторый «словарь» — элементарные смыслы. Однако множество элементарных смыслов само по себе еще не образует языка; язык должен обладать грамматикой — правилами построения предложений из слов. Чтобы выяснить, каким образом строятся предложения на семантическом языке, рассмотрим следующие две фразы: *Это заставляет меня уйти* и *Я вынужден уйти из-за этого*. Очевидно, что они равнозначны; факт их равнозначности подтверждается, в частности, тем, что каждую из этих фраз можно «перевести» предложением *Это — причина моего ухода*, где та же мысль выражена в более явной форме словом *причина*. В составе рассматриваемых фраз имеются некоторые общие части (*я, это, уходить*) и части, которыми они отличаются друг от друга (*заставляет* в противоположность *вынужден*). Кроме того, во втором предложении по сравнению с первым существенным образом изменяется порядок следования частей. Это изменение нельзя считать зависящим от слов, общих для двух предложений (*это, я, уходить*); остается предположить, что оно связано с заменой слова *заставляет* словом *вынужден*. Таким образом, слова *заставляет* и *вынужден* ведут себя в некотором смысле как господствующие: от их выбора зависит порядок остальных слов, которые оказываются в положении подчиненных. Заметим, что слово *заставляет* является господствующим для части *это*, с одной стороны, и части *меня уйти* — с другой. Нетрудно сообразить, что слово *уйти* выполняет ту же роль по отношению к слову *я*.

Из опыта работы с синтаксическими моделями мы уже знаем, что отношение подчинения, устанавливающееся между главным и зависимым элементом, может быть изображено деревом зависимостей; тогда, например, выражение *Это — причина моего ухода* будет иметь следующий вид (см. рис. 17 на стр. 256).

Рассмотрение этого примера позволяет нам сделать второй важный вывод, касающийся синтаксиса семантического языка: в общем случае выражение на семантическом языке имеет вид дерева (или деревьев) зависимости и, следовательно, семантика может описываться как синтаксис. Это представление о строении значений слов является гораздо более глубоким, чем принадлежащая Л. Ельмслеву и уже знакомая нам (см. стр. 61) мысль о том, что

значение слова является простой конъюнкцией «фигур содержания».

Таким образом, семантический язык имеет свои слова (элементарные смыслы) и свою грамматику (деревья зависимостей). Поэтому мы можем обращаться с ним, как с любым другим языком, в частности переводить

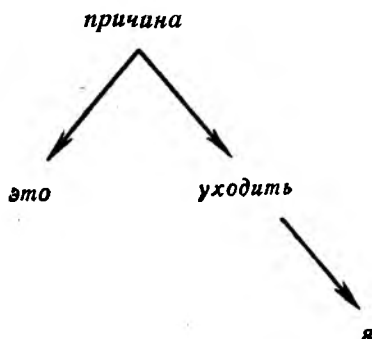


Рис. 17.

тексты с него и на него. Пользуясь понятием перевода, мы можем более точно сформулировать стоящие перед нами задачи. Чтобы смоделировать осмысленное речевое поведение человека, например носителя русского языка, при восприятии текста (анализ), необходимо: 1) иметь русско-семантический словарь и 2) алгоритмы перевода с русского языка на семантический. Чтобы смоделировать осмысленное речевое поведение человека при производстве текста (синтез), необходимо иметь: 1) семантикорусский словарь и 2) алгоритмы перевода с семантического языка на русский.

Ниже мы рассмотрим обе части семантической модели в указанном здесь порядке. Наше изложение будет в основном содержательным, так как в настоящее время формально разработаны и описаны лишь некоторые фрагменты модели, но не вся она целиком. Упор будет сделан на анализ.

Представим себе обычный толковый или двуязычный словарь. Каждая статья в таком словаре имеет две части: определяемая единица, или вход, и ее толкование, или семантическая запись на выходном языке. И выбор мно-

жества определяемых единиц, и выбор множества определяющих единиц, участвующих в толкованиях, выдвигают несколько очень важных вопросов. Мы обсудим сначала вопросы, связанные с выбором множества *определяемых единиц*.

Выше (глава 2, часть II) мы уже говорили о том, что далеко не все имеющиеся в языке слова интересны для лингвиста. Оставаясь в рамках своей науки, лингвист не может точно описать различие между *нейтрино* и *нейтроном*; понятия, необходимые для точного истолкования значений этих терминов, вырабатываются не лингвистом, а физиком, и строгое их определение возможно только на языке физики, но не на языке лингвистики. С этой точки зрения каждый толковый и двуязычный словарь является по существу смешанным: он энциклопедичен ровно в той мере, в какой содержит толкования слов, для определения которых лингвист не располагает необходимыми понятиями.

Русско-семантический словарь строится иным образом: хотя по объему словника он в принципе не должен отличаться от двуязычного словаря, истолкование получают далеко не все входящие в него слова. Без определений остается вся конкретная и терминологическая лексика, в особенности слова, обозначающие конкретные предметы, ср. *стол, иволга, собака, метла, дерево, кастрюля* и т. п. Семантическим кодом каждого такого слова является просто его номер в словаре и, может быть, пояснения типа «мебель», «животное» и т. п. Предметом семантического анализа (разложения на элементарные смыслы) являются слова с более или менее абстрактным значением, обозначающие *отношения* (определение термина «отношение» см. на стр. 114). К их числу относятся глаголы, большая часть прилагательных и наречий и некоторые существительные, такие, как *причина, цель, план, контакт, обладание, знание, ощущение, воля, трудность, выдержка, интуиция, законченность, реальность, желательность, длительность* и т. п. Это, как мы уже знаем (см. стр. 106), слова, значения которых, как можно предполагать, хотя бы в некоторых языках являются грамматическими, т. е. выражаются в обязательном порядке и поэтому входят в структуру кода (правила кодирования сообщений на данном языке). Интересно, что такие слова могут в пределах одного языка превращаться

не только друг в друга, но и в чисто грамматические морфемы, ср. *Петр получил от меня книгу* ↔ *Я дал книгу Петру* (*получить* превращается в *дать*); *Печь задымила* ↔ *Печь начала дымить* (префикс *за-* превращается в слово *начинать*); *Текст переводим* ↔ *Текст может быть переведен* (суффикс *-м* превращается в слово *может*) и т. д. Именно на этой способности слов со значением отношения превращаться друг в друга основаны семантические преобразования, описанные нами в начале данной главы. Обратим внимание читателя и на тот факт, что при всех этих преобразованиях слова, обозначающие предметы, сохраняются.

Анализ последних примеров дает нам основание сделать еще один важный шаг: истолкованию в русско-семантическом словаре подлежат не только значения слов, но и значения грамматических элементов, причем слова и грамматические элементы, имеющие совпадающий семантический код, должны рассматриваться в одной и той же словарной статье. Это относится, в частности, к слову *начинать* и глагольному префиксу *за-*, слову *кончать* и глагольному префиксу *про-* (ср. *он прочитал книгу* ↔ *он закончил чтение книги*), слову *прошлый* и глагольному суффиксу *-л* и т. д.

Другим важным принципом, отличающим русско-семантический словарь от обычных толковых и двуязычных словарей, является переход от описания значений отдельных слов к описанию значений целых высказываний¹: входом словарной статьи в русско-семантическом словаре является образец предложения (или предложений) с данным словом в качестве главного элемента, например *А получает В от С*, или *А зависит от В*, или *А продолжителен*. Этот принцип позволяет по-новому поставить вопрос о синонимии: синонимичными оказываются и те высказывания, которые не содержат синонимов в обычном смысле слова. Более того, обнаруживаются такие семантические связи между словами, которые до сих пор либо вовсе не замечались, либо не были изучены систематически.

Наиболее важным классом связанных по смыслу слов, систематическое изучение которых было впервые начато

¹ Отметим, что аналогичным образом поступали Э. Сэпир <56> и Ш. Балли <242>.

в рамках рассматриваемой семантической модели, являются слова, обозначающие так называемые *о б р а т н ы е*, или *к о н в е р с н ы е*, отношения. Бинарное отношение R^{-1} , обратное к R , определяется следующим образом: $aR^{-1}b$ тогда, и только тогда, когда bRa (93). Примерами выражений, обозначающих обратные отношения, являются выражения *испытывать уверенность* и *внушать уверенность* (ср. *Я не испытываю уверенности в этом* = *Это не внушает мне уверенности*); *бесспорный* и *уверенный* (ср. *Это не кажется мне бесспорным* = *Я не уверен в этом*); *испытывать сомнения* и *вызывать сомнения* (ср. *Я испытываю сомнения в этом* = *Это вызывает у меня сомнения*); *сомнительный* и *сомневаться* (ср. *Это кажется мне сомнительным* = *Я сомневаюсь в этом*). Другими примерами того же типа являются слова *больше* и *меньше*, *раньше* и *позже*; *муж* и *жена*, *строить* и *строиться*, *ненавидеть* и *ненавистен* и т. п., ср. *А больше (позже) В* = *В меньше (раньше) А*, *А — муж В* = *В — жена А*, *Рабочие строят дом* = *Дом строится рабочими*, *Он ненавидит пошлость* = *Пошлость ненавистна ему*. Из этих примеров следует, между прочим, что регулярным грамматическим средством выражения обратных отношений является *п а с с и в и з а ц и я*.

Поскольку высказывания, образованные выражениями, которые обозначают обратные отношения, *с и н о н и м и ч н ы* друг другу, т. е. переводятся одним и тем же предложением семантического языка, они могут толковаться в одном месте (словарной статье) русско-семантического словаря. Здесь обнаруживается интереснейшая особенность русско-семантического словаря, который объединяет в себе некоторые существенные черты толковых и тематических (или «идеологических» словарей). В толковых словарях описываемые единицы (слова) располагаются в алфавитном порядке, и каждое слово снабжается толкованием. Поскольку для группировки слов используется столь случайный, хотя и удобный, признак, как номера букв слова в алфавите, одно и то же значение толкуется в словаре столько раз, сколько различных способов выражения оно имеет в данном языке. В отличие от этого в тематическом словаре слова сгруппированы по семантическому признаку: в одну словарную статью попадают все слова и выражения, которые имеют *о б щ и е* семантические признаки. Однако в тематическом словаре

отсутствуют определения значений. Русско-семантический словарь рассматриваемого здесь типа сходен с тематическим словарем в том отношении, что в рамках одной словарной статьи объединяет все синонимичные высказывания (вернее, образцы синонимичных высказываний), и сходен с толковым словарем в том отношении, что содержит определения значений: входу словарной статьи, т. е. множеству синонимичных образцов высказываний, сопоставляется образец предложения на семантическом языке, который является их переводом.

Рассмотрим теперь вопросы, связанные с выбором множества определяющих единиц. В толковых словарях не проводится различия между определяемыми и определяющими словами; считается, что и те и другие принадлежат одному и тому же языку, например русскому в случае толковых словарей русского языка. Поэтому в принципе и в левой (определяемой) и в правой (определяющей) части словарной статьи могут встречаться одни и те же слова. Это и является источником столь обычных в толковых словарях тавтологических (круговых) определений типа *поддержка* = «помощь, содействие»; *помощь* = «содействие кому-нибудь в чем-нибудь»; *содействие* = «помощь, поддержка в какой-нибудь деятельности» или *большой* = «значительный по размерам, по величине, силе»; *значительный* = «большой по размерам, силе» <128>.

В отличие от этого в русско-семантическом словаре четко различаются определяемые единицы, т. е. выражения естественного языка, и определяющие единицы, т. е. слова семантического языка (= элементарные смыслы). Это принципиальное различие между описываемым языком, который иначе называется языком-объектом, и описывающим языком, который часто называют языком-субъектом, или метаязыком, сохраняется и в том случае, когда по каким-либо причинам внешнего характера толкование значения фактически записывается словами того же языка. Оно сохраняется потому, что все толкования имеют совершенно определенную логическую структуру: они сводимы к очень небольшому набору простейших неопределяемых понятий, заимствованных из математики или столь очевидных, что они не нуждаются в пояснениях. Это такие понятия, как *множество*, *предмет*, *свойство*, *отношение*, *необходимость*, *достаточность*, *истинность*, *не*, *и*, *или*,

всякий, один, два, количество, время, пространство и некоторые другие.

На основе этих исходных неопределяемых понятий определяются промежуточные понятия типа *никакой, другой, только, несовместимо, предшествовать, соприкасаться, каузировать, продолжительность* и т. п.

В определениях значений могут участвовать неопределяемые понятия, промежуточные понятия и любые уже определенные понятия; при этом следует помнить, что использование промежуточных и уже определенных понятий является техническим приемом сокращения определений; любое определение в случае необходимости может быть переписано в более развернутой форме, составленной исключительно из выражений семантического языка.

Считается, что значением каждого определяемого выражения является внеязыковая ситуация, которая и изображается семантической записью. Определение для того или иного выражения подыскивается в ходе интуитивных рассуждений, которые носят название *метода портретирования*. Этот метод состоит в последовательном описании «картинок», которые в совокупности составляют типовую ситуацию, обозначаемую данным выражением. Так, значение слова *цель* нащупывается следующими содержательными рассуждениями. *Цель* — это положение вещей, являющееся *желательным* для некоторого лица *А*. *Цель* отличается от *мечты* наличием реальных *путей* к ее осуществлению, а от прочих *желаний* — тем, что лицо *А* само ее осуществляет, используя при этом имеющиеся в его распоряжении *ресурсы*. Для достижения цели лицо *А* *действует* или *планирует* действия, которые кажутся ему *целесообразными*, и т. д. Когда значение слова таким образом нащупано, оно получает более строгое определение на описанном выше семантическом языке. Подчеркнем, что правильность найденного определения подтверждается не объективностью процедуры, с помощью которой оно отыскивается, а другими средствами, например эффективностью автоматического перевода, осуществляемого на основе данной семантической модели.

Благодаря отчетливой формулировке определений становятся возможными нетривиальные сопоставления сходных, но не совпадающих по значению выражений естественного языка. Ниже мы приведем, без особых комментариев, несколько примеров такого рода, заимствованных

нами из «словарной» части рассматриваемой семантической модели.

- I. *А пренебрегает В* (ср. *этим различием можно пренебречь*) = *А* исходит из того, что роль *В* в данном множестве фактов очень мала
А игнорирует В (ср. *он игнорирует меня*) = Желательность для *А* считать роль *В* малой каузирует (заставляет) *А* пренебрегать *В*
- II. *А старается (сделать) Р* (ср. *Он старался удержаться на ногах*) = *А* по плану *С* прилагает много или все ресурсы, нужные для *Р*
А пытается (сделать) Р (ср. *Он пытается решить задачу*) = *А* старается сделать *Р*, не имея истинного мнения о том, является ли его план *С* путем к *Р*
А пробует (сделать) Р (ср. *Он пробует решить вопрос по-своему*) = *А* пытается сделать *Р*; *А* считает, что его попытка каузирует либо *Р*, либо истинное мнение, что *Р* не годен
- III. *Лицу А удалось Р; Р—удача А* (ср. *Ему удалось пройти первым*) = *А* прилагало ресурсы для *Р*; *Р* каузировано; роль *А*—не единственный фактор, каузировавший *Р*
Лицо А добилось Р; лицо А достигло Р; Р—достижение А (ср. *Он добился признания*) = Лицо *А* приложило много ресурсов, которые каузировали *Р*—цель лица *А*
Лицу А повезло с Р; лицу А посчастливилось с Р (ср. *Ему повезло с оппонентами*) = Некоторый *Х* каузировал *д*, нужное для *Р*—цели *А*; *Х* не зависит от *А*

На этом мы заканчиваем обсуждение вопросов, связанных со словарной частью семантической модели, и переходим ко второй ее части, т. е. алгоритмам перевода с естественного языка на семантический. В настоящее время существуют только пробные алгоритмы такого рода; поэтому мы не будем входить в какие-либо технические детали, а сконцентрируем внимание на самой проблеме.

Переводу на семантический язык подлежит предложение естественного языка с уже найденным деревом синтаксических зависимостей между словоформами. Перевод проводится в три этапа. На первом этапе каждый куст дерева синтаксических зависимостей, содержащий слово, которое обозначает отношение, и все зависимые от него слова, переписывается с помощью русско-семантического словаря в виде семантического дерева зависимостей.

Мы рассмотрим в качестве иллюстрации упрощенный пример, с помощью которого можно пояснить идейную сторону дела, но отнюдь не фактически применяемую технику. Пусть, например, в составе какого-то сложноподчиненного предложения имеется простое предложение *Мать лишает сына завтрака*. Соответствующий ему куст синтаксического дерева имеет следующий вид (см рис. 18).

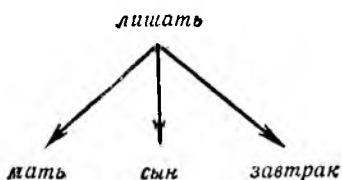


Рис. 18.

Как мы помним, *лишить* значит «заставить перестать иметь». Поэтому дерево семантических зависимостей для этого куста должно принять следующий вид (см. рис. 19).

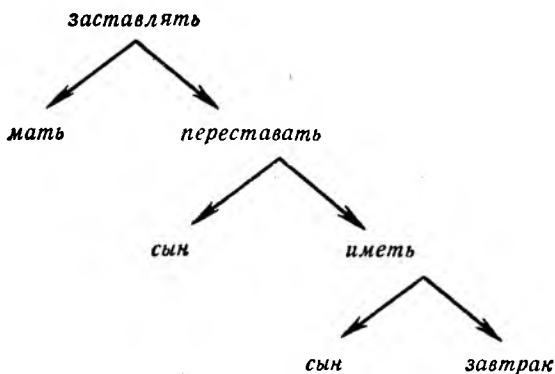


Рис. 19.

На втором этапе семантические деревья, построенные аналогичным образом для каждого из кустов синтаксического дерева, сращиваются, по определенным правилам,

в семантическое дерево всего предложения. Следовательно, на втором этапе все предложение целиком переводится на семантический язык. На третьем этапе полученная запись приводится к стандартному виду с помощью семантических равенств. Простейший пример семантического равенства — правило снятия двойного отрицания, которое равносильно утверждению.

Модели синтеза осмысленных предложений основаны на том же круге идей, и поэтому мы не будем их специально рассматривать, тем более что разработаны они гораздо менее подробно, чем соответствующие модели анализа. Некоторые не лишённые интереса детали читатель найдет в <57> и <97>.

* * *

На этом мы заканчиваем рассмотрение моделей речевой деятельности человека. Как помнит читатель, при построении моделей этого типа предполагается заранее известной вся информация о данном языке, включая информацию о значениях его грамматических форм и слов. Задача лингвиста заключается не в том, чтобы объективными методами обосновать выбор используемых понятий (хотя, вообще говоря, неплохо, если он может это сделать), а в том, чтобы формализовать их для объяснения процессов понимания и производства текстов человеком, который владеет данным языком (т. е. знает его грамматику и значения слов).

Если единственной гарантией правильности модели исследования является объективность процедур, с помощью которых лингвист получает интересующие его сведения о языке, то правильность модели речевой деятельности можно проверить и другими способами, экспериментальными и теоретическими. О некоторых из них мы расскажем в последней части нашей книги.

ЧАСТЬ V

ПОНЯТИЕ МЕТАТЕОРИИ

Рассмотрим следующую модель (алгоритм) порождения текстов, реализованную на электронной вычислительной машине Манчестерского университета. В качестве исходной информации в ней используется связный текст, достаточно большой для того, чтобы каждая словоформа была употреблена в нем минимум дважды. В число словоформ включаются и точечные знаки (точка). Правила образования новых текстов — следующие.

В тексте произвольно выбирается некоторая словоформа, стоящая непосредственно после точечного знака (т. е. начальное слово произвольного предложения). Она выписывается. Затем в данном тексте отыскивается второй случай употребления той же словоформы и находится словоформа, стоящая справа от нее. Эта вторая словоформа выписывается справа от уже выписанной словоформы. Отыскивается второй случай употребления второй словоформы и находится словоформа, стоящая справа от нее. Третья словоформа приписывается справа к двум ранее выписанным и т. д., пока не встретится точечный знак. По этим правилам был синтезирован следующий текст: *«Мое маленькое сокровище! Моя вразумительная привязанность чудесно привлекает твой ласковый восторг. Ты мое любящее обожание, мое распирающее грудь обожание. Мое братское чувство с затаенным дыханием ожидает твоего дорогого нетерпения. Обожание моей любви нежно хранит твой алчный пыл. Твой тоскующий МУК»* <140, 165>.

Несмотря на внешне неплохие результаты, модель не вызывает у нас особого доверия: слишком примитивны используемые ею правила и слишком сложна исходная

информация (модель строит новые тексты не из более элементарных частей, а из текстов же).

Перед нами возникает вопрос о том, является ли данная модель единственной моделью порождения текста, а если она не является таковой, то как выбрать лучшую.

На первый вопрос ответить нетрудно. Из того, что было сказано в главе I части II о лингвистической модели как функциональной аппроксимации объекта, следует, что данное языковое явление может быть объяснено более чем одной моделью <19>, <384>, <306>. Разные модели одного и того же явления могут отличаться друг от друга следующим:

1. Характером исходной информации. Так, фонологические системы естественных языков могут моделироваться не только на уровне фонем, но и на уровне дифференциальных признаков фонем. Об этом свидетельствуют экспериментальные и теоретические работы Р. Якобсона и его школы, кратко охарактеризованные нами на стр. 70.

2. Объемом исходной информации. Примером могут служить три типа исследовательских моделей, которые мы назвали на стр. 101: модели первого типа используют в качестве входной информации только текст, модели второго типа — текст и множество правильных фраз данного языка, а модели третьего типа — текст, множество правильных фраз и множество семантических инвариантов. Естественно, что использование более богатой входной информации позволяет, при прочих равных условиях, получить более богатую информацию на выходе, и, следовательно, модели такого рода будут неизбежно отличаться друг от друга и по степени аппроксимации объекта.

3. Принципами обработки информации, или характером используемых правил. Мы видели выше, что правила могут быть вероятностными и строгими, а среди строгих могут выделяться дистрибутивные, трансформационные и иные типы правил.

4. Формой изложения. Выше мы довольно подробно обсудили две наиболее интересные формы изложения модели — исчисление и алгоритм. Ими, конечно, не исчерпываются существующие в этой области возможности.

Поскольку в общем случае некоторое явление может быть смоделировано более чем одним способом, мы должны располагать системой оценок, с помощью которой мы могли бы сравнивать различные модели одного и того же явления между собой и выбирать ту из них, которая является наилучшей аппроксимацией объекта или лучше других моделей приспособлена для решения данной конкретной задачи. Назначение метатеории и состоит в том, чтобы обеспечить лингвистическую теорию такой системой оценок.

Впервые систему критериев для оценки лингвистической теории предложил Л. Ельмслев; провозглашенный им и уже упоминавшийся нами принцип эмпиризма гласит: «Описание должно быть свободным от противоречий... исчерпывающим и предельно простым» <53, 272>. В разное время к этим трем критериям последовательности, полноты и простоты добавлялись критерии адекватности <199>, <297>, красоты <279> и симметрии <300>; последние два совпадают, по всей видимости, с критерием простоты <286>.

Из названных здесь понятий критерий последовательности не нуждается в специальном определении, так как он и без того очевиден. Что касается понятий полноты, адекватности, простоты и т. п., то они не являются достаточно очевидными, чтобы их использовать без определений. В связи с этим мы изложим принадлежащую И. А. Мельчуку попытку формализовать, на базе элементарных представлений из области теории множеств, эти и некоторые другие критерии оценки качества лингвистических моделей <115>, <117>¹. Заметим, что предложенные И. А. Мельчуком способы измерения простоты, полноты и других интересных свойств моделей никем не были еще практически опробованы отчасти потому, что в современной лингвистике нет достаточного числа моделей, записанных в требуемой для такого рода оценки стандартной форме, а отчасти и потому, что критерии И. А. Мельчука носят в значительной мере предварительный характер. Следовательно, мы излагаем эту работу не столько для того, чтобы вооружить читателя средством проверки

¹ См. также работу Ф. Харвуда <297>, формализовавшего понятия полноты и адекватности модели.

качества лингвистических теорий, сколько для того, чтобы продемонстрировать принципиальную возможность строго говорить о категориях, которые на первый взгляд не поддаются формализации.

Для обсуждения этого вопроса нам понадобятся некоторые понятия теории множеств, введенные в части II. Напомним, что мощностью множества M (обозначение — $|M|$) называется число входящих в него элементов, а пересечением двух множеств M и N называется такое множество $M \cap N$, элементы которого принадлежат одновременно и M , и N .

Перейдем к существу предложений И. А. Мельчука. Обозначим буквой M множество элементов, образующих то языковое явление, которое нас интересует (например, множество форм глагола). Буквой N обозначим множество элементов (множество глагольных форм), порождаемых моделью. Буквой b обозначим число (не множество!) исходных неопределяемых понятий, используемых в модели; наконец, буквой c обозначим число правил, по которым из исходных элементов образуются конечные объекты. Число b показывает абсолютную эконо-
мичность модели: чем оно меньше, тем экономичнее модель. Число c показывает абсолютную простоту модели: чем меньше правил, тем проще модель.

Полнота модели измеряется формулой

$$P = \frac{|M \cap N|}{|M|}.$$

Чтобы пояснить содержательный смысл этой функции, рассмотрим два крайних случая (см. рис. 20).

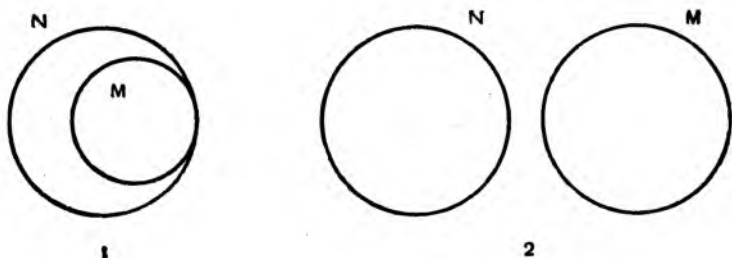


Рис. 20.

В первом из них модель порождает все множество глагольных форм: M включено в N . Это значит, что пересечение $M \cap N$ равно M и, следовательно, в данном случае

$$P = \frac{|M \cap N|}{|M|} = \frac{|M|}{|M|} = 1.$$

Во втором случае модель не порождает ни одной глагольной формы: пересечение $M \cap N$ пусто. Следовательно, мощность пересечения равна нулю и

$$P = \frac{|M \cap N|}{|M|} = \frac{0}{|M|} = 0.$$

И. А. Мельчук предложил пользоваться и более тонким критерием полноты модели. Допустим, что нам требуется оценить полноту модели, порождающей глагольные словосочетания с существительным в творительном падеже. Допустим, кроме того, что она порождает словосочетания наиболее распространенных типов (*качать головой, пахать плугом, восхищаться книгой, лететь стрелой, идти лесом, знать ребенком* и т. п.), но не порождает устаревших и поэтому чрезвычайно редких словосочетаний с творительным причины (ср. *Осел мой глупостью в пословицу вошел*). В данном случае более тонкую оценку полноты модели даст такая функция, которая учитывает не только число порождаемых типов словосочетаний, но и частоту каждого типа в тексте. Если гипотетическая модель, которую мы здесь рассмотрели, будет оцениваться с помощью более тонкой функции, она окажется достаточно полной, несмотря на то, что не порождает словосочетаний определенного типа.

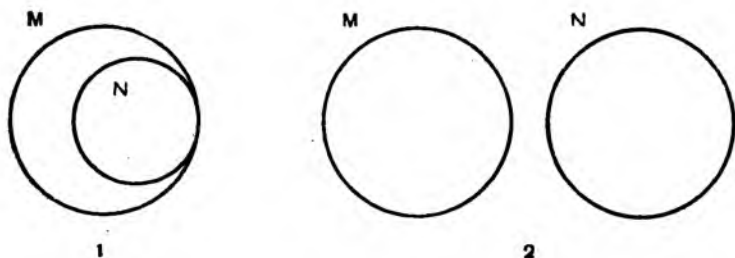


Рис. 21.

Адекватность модели измеряется формулой

$$A = \frac{|M \cap N|}{|N|}.$$

Чтобы пояснить содержательный смысл этого понятия, мы рассмотрим два крайних случая (рис. 21).

В первом из них модель не порождает ни одной «лишней» формы, не принадлежащей множеству M : N включено в M . Это значит, что пересечение $M \cap N$ равно N и, следовательно,

$$A = \frac{|M \cap N|}{|N|} = \frac{|N|}{|N|} = 1.$$

Во втором случае модель порождает только «лишние» формы, ни одна из которых не принадлежит M : пересечение $M \cap N$ пусто. Следовательно, мощность пересечения равна нулю и

$$A = \frac{|M \cap N|}{|N|} = \frac{0}{|N|} = 0.$$

Из вида функций ясно, что модель может быть полной, но не вполне адекватной (ср. рис. 20) или адекватной, но недостаточно полной (ср. рис. 21); модель является и полной и адекватной, когда порожаемое множество N в точности совпадает с реальным множеством M , т. е. когда она порождает все объекты данной совокупности, и только такие объекты.

Экономичность модели измеряется формулой

$$E = \frac{|N| + 1}{|N| + b}.$$

Число неопределяемых понятий, используемых в модели, не может быть меньше одного; с этим идеальным, «эталонным», случаем и сравнивается реальная модель, содержащая b неопределяемых понятий. Экономичность модели максимальна (равна 1), когда $b=1$; чем больше b , тем меньше, при неизменном $|N|$, экономичность модели.

Простота модели измеряется формулой

$$S = \frac{|N| + 1}{|N| + c}.$$

Число правил, по которым из исходных элементов строятся конечные объекты, также не может быть меньше

одного; с этим идеальным случаем сравнивается реальная модель, содержащая s таких правил. Простота модели максимальна (равна 1), когда $s=1$; чем больше правил, тем, при неизменном $|N|$, сложнее модель.

Как ясно из вида двух последних формул, для измерения экономичности и простоты модели недостаточно знать абсолютное число исходных понятий и правил над ними; если, например, две модели используют набор из 5 исходных понятий и двух правил, это еще не значит, что они в одинаковой мере просты и экономичны. Все зависит от того, какова мощность порождаемого ими множества объектов. Легко проверить, что более экономичной и простой, при одинаковом числе исходных понятий и правил над ними, окажется та модель, которая порождает более мощное множество объектов.

Иногда между экономичностью и простотой существует обратная зависимость: увеличение числа исходных понятий позволяет сократить число правил, и наоборот. В таких случаях, в зависимости от конкретных условий той или иной задачи, мы можем выбрать более экономичную, хотя и менее простую модель (или наоборот). Так, если электронная вычислительная машина, осуществляющая синтез глагольных форм, имеет память очень значительного объема, но скромные логические возможности, разумно выбрать простой алгоритм синтеза, даже если он не очень экономичен; если же машина обладает небольшой памятью, но богатыми логическими возможностями, для нее более выгодным будет по возможности экономичный алгоритм, даже если он не очень прост.

Обычно полнота и адекватность рассматриваются как признаки внешней оправданности теории, т. е. ее согласия с опытными данными, а экономичность и простота — как признаки ее внутреннего совершенства $\langle 53 \rangle$, $\langle 199 \rangle$, $\langle 286 \rangle$. Следует, однако, иметь в виду, что всякая внутренне совершенная теория имеет не только эстетические преимущества по сравнению с теорией, лишенной внутреннего совершенства, но и серьезные преимущества по существу. Поскольку она исходит из небольшого числа простых и общих предпосылок и не содержит произвольных предположений, не вытекающих из них однозначным образом, она глубже всего способна объяснить реальное единство своего объекта. Поэтому очень часто поиск простой и экономичной теории

оборачивается поиском истинного знания, а принятие сложной и неэкономичной теории означает примирение с непринципиальным и недолговечным квазиобъяснением. Интересно, что к теории ларингальных (см. стр. 94), блистательно подтвержденной позднейшими открытиями, Ф. де Соссюра привел поиск простого решения вопроса об индоевропейской системе гласных и нежелание довольствоваться сложными построениями современной ему науки.

Рассмотренные выше критерии создают основу для экспериментальной оценки и сравнения моделей. Чтобы из нескольких моделей выбрать лучшую, необходимо экспериментальным (например, машинным) путем получить результаты работы каждой из них и вывести для них числовые оценки по указанным выше формулам. Чтобы убедиться в негодности некоторой модели, необходимо все-таки не только построить ее от начала до конца, но и обработать с ее помощью некоторый материал. Естественно возникает вопрос о том, существует ли метод оценки совершенства моделей, не связанный с проведением громоздких экспериментов. Оказывается, что такой метод существует: он состоит в частичной замене экспериментов теоретическими доказательствами. Здесь мы подходим к самому увлекательному разделу метатеории и одному из интереснейших вопросов современной лингвистики — вопросу о теории грамматик.

Если, например, нам предъявлены две модели, описывающие один и тот же материал, причем свойства моделей заданы в явном виде, и если, кроме того, известны существенные свойства материала, то в принципе мы можем доказать теоремы, из которых будет следовать, что одна из моделей в некоторых отношениях более совершенна, чем другая. Более того, для доказательства таких теорем, вообще говоря, не нужно располагать готовыми моделями; достаточно знать только существенные свойства некоторого класса моделей, и тогда все, что доказано для моделей данного класса вообще, будет, конечно, верно и для каждой отдельной модели.

Специфика этой книги не позволит нам сколько-нибудь полно рассказать об этой области современной структурной лингвистики, разработанной прежде всего Н. Хомским <199>, <202>, а после него Е. Бахом <240>, А. В. Глад-

ким <43а>, О. С. Кулагиной <88>, И. И. Ревзиным <152>, С. Я. Фитиаловым <190>, Г. С. Цейтиным и другими исследователями, и мы вынуждены будем ограничиться несколькими несложными разъяснениями. Мы надеемся, что читателя не отпугнет сухость и абстрактность нижеследующих рассуждений; такова специфика предмета, о котором мы будем говорить.

Мы уже сказали, что для доказательства интересных лингвистических теорем надо знать некоторые свойства моделируемого объекта и самих моделей. Мы начнем с обсуждения первого вопроса.

Определим язык как множество предложений из конечного алфавита символов. Как мы уже знаем (стр. 90), в качестве алфавита символов может быть взято либо множество фонем, либо множество морфем, либо множество словоформ, либо, наконец, множество символов, обозначающих классы морфем или классы словоформ. Предложения на таком языке имеют вид цепочек фонем (или морфем, или словоформ, или символов классов). Язык задается указанием его алфавита и правил, по которым из элементов алфавита строятся предложения.

Рассмотрим следующие два искусственных языка с алфавитом символов a и b , содержащих всего по одному правилу построения предложений <199>:

(1) $aa, bb, abba, baab, aaaa, bbbb, aabbaa$ и вообще все предложения, состоящие из цепочки X , за которой следует «зеркальное отражение» X (т. е. X в обратном порядке), и только такие предложения;

(2) $aa, bb, abab, baba, aaaa, bbbb, aabaab$ и вообще все предложения, состоящие из цепочки X (содержащей в некоторой комбинации буквы a и b), за которой следует точно такая же цепочка X , и только такие предложения.

Какими бы примитивными ни казались эти языки, в некотором отношении они представляют исключительный интерес: на их «материале» мы можем в чистом виде выделить и изучить два весьма важных свойства некоторых предложений или фрагментов предложений естественных языков, а именно свойства прямого и зеркального повторения элементов предложения.

Зеркальность проявляет себя тем, что для любого элемента первой половины цепочки во второй ее половине найдется ровно один зависимый от него элемент. В качестве

естественного примера зеркальности рассмотрим предложения типа *Если теория неприменима, то надо искать другое решение*, предложения типа *Либо теорема недоказуема, либо неприменима* и предложения типа *Теорема, которую вы сформулировали, недоказуема*. В каждом из этих предложений имеется пара зависимых друг от друга элементов: (а) *если . . . , то* (слово *то* в предложениях такого типа нельзя заменить словом *либо*, или *что*, или *потому*; после того как в первой половине предложения появилось слово *если*, во второй его половине обязательно должно появиться зависящее от него слово *то* (или *тогда*); (б) *либо . . . либо* (второе *либо* зависит от первого и должно встретиться во второй половине любого сложного предложения, открывающегося словом *либо*); (с) *теорема . . . недоказуема* (грамматические категории слова *недоказуема* — ж. р., ед. ч., им. п. — повторяют, по правилам согласования, грамматические категории слова *теорема* и не могут быть заменены никакими другими категориями, если одновременно не изменится слово *теорема*).

Если допустить, как это было сделано в главе I части II, что длина предложения естественного языка в принципе ничем не ограничена, то можно представить себе следующий процесс построения сложного предложения из предложений указанного выше типа (символы S_1, S_2, \dots обозначают предложения):

Если S_1 , то S_2 .

Пусть S_1 — предложение типа (б): *либо S_3 , либо S_4* . Тогда возможно предложение

Если, либо S_3 , либо S_4 , то S_2 .

Пусть, наконец, S_3 — предложение типа (с): *теорема, S_5 , недоказуема*. Тогда возможно предложение

Если, либо теорема, S_5 , недоказуема, либо S_4 , то S_2 .

Подставляем вместо символов S_5, S_4 и S_2 соответствующие реальные предложения из приведенных выше примеров и получаем *Если, либо теорема, которую вы сформулировали, недоказуема, либо неприменима, то надо искать другое решение*.

Легко заметить, что часть этого предложения, вполне правильного, хотя и несколько громоздкого, обладает свойством зеркальности (ср. цепочку элементов *если + либо + ж. р., ед. ч., им. п.* в первой части предложе-

ния и обратную ей цепочку элементов *ж. р., ед. ч., им. п. + + либо +то* во второй части предложения). Если в каждом естественном языке найдется хотя бы по одному примеру такого рода, то это явится достаточным доказательством того, что в общем случае предложения на естественных языках обладают свойством зеркальности.

Возможны в естественных языках и цепочки типа (2) с прямым повторением элементов предложения. Самым обычным примером этого типа являются сложноподчиненные предложения (дополнительные, определительные и др.) с последовательным подчинением, ср. *Он сказал, что S_1 , что S_2 , что S_3 , . . . , что S_i (Он сказал, что он думает, что все знают, что он нашел решение задачи) или Вот человек, который S_1 , который S_2 , который S_3 , . . . , который S_n (Вот пес без хвоста, который за шиворот треплет кота, который пугает и ловит синицу, которая часто ворует пшеницу, которая в темном чулане хранится в доме, который построил Джек).*

Итак, мы указали некоторые интересные свойства естественных языков. У них, конечно, имеются и другие интересные свойства (см. указанные выше работы), в том числе и свойства, никогда не анализировавшиеся классической грамматикой (см., например, стр. 248). Однако для наших целей достаточно приведенного здесь материала, и мы перейдем ко второму вопросу — рассмотрению некоторых свойств грамматик, моделирующих те или иные языковые явления. Речь пойдет, в основном, о так называемой грамматике с конечным числом состояний и — в меньшей мере — о грамматике непосредственно составляющих, как они изложены Н. Хомским <199>. Другие типы грамматик будут лишь вскользь упомянуты.

Представим себе порождающее устройство, которое может принимать конечное число внутренних состояний. Неважно, какую картину этого устройства читатель нарисует в своем воображении; на худой конец можно представить себе приемник, регулятор которого может быть выключен или включен и настроен на короткие, средние или длинные волны. Различные состояния порождающего устройства обозначим символами $C_0, C_1, C_2, \dots, C_n$. При переходе из одного состояния в другое устройство выдает некоторый языковой объект (например, фонему, или букву, или морфему, или словоформу, или символ класса словоформ, или пробел между словоформами и т. п.):

$C_i C_j \rightarrow a_{ij}$. При переходе из некоторого состояния C_i в начальное состояние C_0 выдается точечный знак. Последовательность состояний, проходимых устройством за один прием (цикл), соответствует некоторой фразе. Множество фраз, порожденных таким устройством, образует язык с конечным числом состояний, а грамматика, описывающая такой язык, называется грамматикой с конечным числом состояний.

Рассмотрим «язык», состоящий из следующего двустий В. Хлебникова (пример заимствован у И. И. Ревзина (152, 118)):

*Где качались тихо ели,
Где шумели звонко ели...*

Грамматика этого, как и любого другого, «языка» может быть задана тривиальным образом — простым перечислением всех правильных фраз данного языка (в списке таких фраз содержатся, правда, имплицитно, все грамматические правила языка, и человек, выучивший наизусть все фразы списка, сумеет построить другие фразы на том же языке). Разумно, однако, попытаться построить грамматику даже этого простого «языка» менее тривиальным и более экономным образом. Нетривиальной и более экономной грамматикой нашего «языка» является следующая грамматика с конечным числом состояний:

Алфавит символов: *где, качались, шумели, тихо, звонко, ели.*

Состояния: $C_0, C_1, C_2, C_3, C_4, C_5$.

Грамматические правила:

$C_0 C_1 \rightarrow \text{где}$	$C_2 C_4 \rightarrow \text{тихо}$
$C_1 C_2 \rightarrow \text{качались}$	$C_3 C_4 \rightarrow \text{звонко}$
$C_1 C_3 \rightarrow \text{шумели}$	$C_4 C_5 \rightarrow \text{ели}$

Порождение грамматических фраз данного «языка» по правилам сформулированной выше грамматики можно представить следующей диаграммой (см. рис. 22).

Чтобы грамматика такого типа могла порождать предложения любой наперед заданной длины и чтобы число предложений было бесконечно, достаточно добавить в нее некоторые рекурсивные правила, разрешающие возвращаться в уже пройденные состояния; на диаграмме такие правила имеют вид обратных стрелок (см. рис. 23).

Усовершенствованная таким образом грамматика породит, в частности, предложение *Где качались шумели звонко ели.*

Наконец, каждому переходу из одного состояния в другое может быть приписана некоторая вероятность, и модель языка с конечным числом состояний будет завершена ¹.

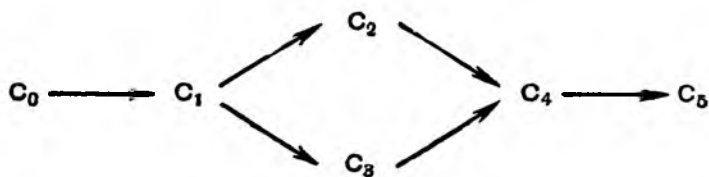


Рис. 22.

Перейдем к грамматике непосредственно составляющих ². Грамматика непосредственно составляющих определяется, как мы помним, алфавитом символов, конечным набором цепочек $S_1, S_2, S_3, \dots, S_i$ (типов предложений) и конечным упорядоченным набором

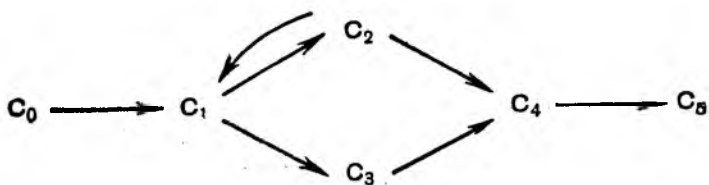


Рис. 23.

правил подстановки f_1, f_2, \dots, f_j (правил образования предложений). Рассмотрим следующий абстрактный пример такой грамматики. Алфавит символов — a, b . Набор цепочек — S . Правила подстановки — (1) $S \rightarrow ab$ и (2)

¹ Интересные эксперименты с моделью языка с конечным числом состояний были проведены за рубежом и в Советском Союзе. См., в частности, эксперимент, описанный в работе И. М. Яглома, Р. Л. Добрушина и А. М. Яглома <234>.

² Подробно порождающие грамматики непосредственно составляющих рассмотрены нами в IV части книги; в III части рассмотрена экспериментальная модель непосредственно составляющих.

$S \rightarrow aSb$. Правильными предложениями, порождаемыми этой грамматикой, являются предложения вида ab , $aabb$, $aaabbb$, $aaaabbbb$ и т. п. Последнее предложение получается в результате трехкратного применения второго правила и однократного применения первого правила:

- | | |
|-------------------------|------------|
| (1) $S \rightarrow aSb$ | aSb |
| (2) $S \rightarrow aSb$ | $aaSbb$ |
| (3) $S \rightarrow aSb$ | $aaaSbbb$ |
| (4) $S \rightarrow ab$ | $aaaabbbb$ |

Последовательность цепочек, в которой каждая следующая цепочка S_j получается в результате применения к предыдущей цепочке S_i некоторого правила подстановки f_k , называется **выводом**. Вывод считается **завершенным**, если нет такого правила f_j , с помощью которого можно было бы преобразовать его последнюю цепочку. Вывод (4) **завершен**, а вывод (3) **не завершен**. Последняя цепочка **завершенного вывода** называется **терминальной**; к ней не применимы никакие правила подстановки из числа имеющихся. В нашем примере цепочка $aaaabbbb$ является терминальной, а предыдущая цепочка $aaaSbbb$ таковой не является, так как она содержит символ S , к которому применимы оба имеющиеся у нас правила подстановки. Множество терминальных цепочек, порожденных такой грамматикой, образует **терминальный язык**.

Итак, мы изучили некоторые свойства синтаксической структуры предложений естественных языков (зеркальное и прямое повторение структурных элементов предложения), два типа языков (языки с конечным числом состояний и терминальные языки) и два типа грамматик (грамматики с конечным числом состояний и грамматики непосредственно составляющих). Теперь мы можем связать теоремами изученные нами свойства языков и грамматик и сравнить разные типы грамматик с точки зрения их адекватности, не прибегая к экспериментам и даже не строя до конца этих грамматик. Мы не будем приводить здесь никаких формальных доказательств, но укажем путь, на котором они могут быть получены. Формальные доказательства даны в <202>.

Теорема 1 (Н. Хомский). *Всякий язык с конечным числом состояний является терминальным языком; обратное неверно: существуют терминальные языки, которые не являются языками с конечным числом состояний.*

Эту теорему можно вывести как следствие из других теорем, гласящих, что язык с конечным числом состояний не содержит зеркальных цепочек и, следовательно, грамматика с конечным числом состояний их не порождает, в то время как терминальный язык в общем случае содержит такие цепочки, а грамматика непосредственных составляющих может их порождать.

Теорема 2. *Естественные языки не являются языками с конечным числом состояний.*

Это следует из того, что в них возможны предложения с зеркальным и прямым повторением структурных элементов, которые не порождаются грамматиками с конечным числом состояний. Следовательно, грамматики с конечным числом состояний не адекватны структуре естественных языков (не порождают всех возможных типов предложений). В этом смысле грамматики непосредственных составляющих являются более совершенными, поскольку в них легко встраиваются механизмы порождения предложений «трудных» типов. Однако даже они не вполне адекватны структуре естественных языков, так как в естественных языках существуют такие типы предложений, которые они не в состоянии породить. К ним, в частности, относятся рассмотренные выше предложения с прямым повторением элементов, которые, как показал Н. Хомский, могут быть порождены только более сильной трансформационной грамматикой.

В последние годы были доказаны некоторые другие важные и интересные теоремы, в частности теорема об эквивалентности (в определенных границах) двух весьма распространенных способов записи синтаксической структуры предложения: дерева непосредственных составляющих и дерева зависимостей <98>, <133>.

Эти и им подобные теоретические доказательства являются необходимой частью лингвистической теории, хотя, конечно, они не устраняют нужды в широких экспериментальных исследованиях, и особенно решающих экспериментах.

ЗАКЛЮЧЕНИЕ

Науку часто сравнивают с кораблем, который время от времени перестраивается сверху донизу, оставаясь все время на плаву. Перестройка происходит всякий раз, когда накоплены факты, для понимания которых недостаточно старых объяснений. В такие периоды наука переключает часть своих сил со сбора и классификации новых фактов на теоретическое осмысление уже накопленного материала. Именно такую эпоху переживает сейчас лингвистика: возникновение современной структурной лингвистики и было выражением поворота от сбора новых фактов к теоретическому осмыслению уже накопленных фактов. У некоторых языковедов, не представляющих себе научной работы без обследования новых материалов, теоретические исследования, не вводящие в обращение ни одного нового факта, вызывают недоумение и протест. Сошлемся поэтому на пример и авторитет величайшего ученого современности А. Эйнштейна, который никогда не был экспериментатором и открытия которого выросли из размышлений над хорошо известными фактами. «Постановка новых вопросов,— писал А. Эйнштейн,— развитие новых возможностей, рассмотрение старых проблем под новым углом зрения требует творческого воображения и отражает действительный успех в науке. Принцип инерции, закон сохранения энергии были получены только благодаря новым и оригинальным идеям в отношении уже хорошо известных экспериментов и явлений» (231, 99).

Ошибется тот, кто будет оценивать достижения структурной лингвистики только объемом новых языковых материалов, введенных ею в оборот. Своими наиболее важными и существенными результатами она обязана применению новых идей к уже известным фактам.

Развитие структурной лингвистики привело прежде всего к обогащению и углублению традиционной лингвистической проблематики. Арсенал технических приемов лингвистики пополнился принципами дистрибутивного, глоссематического и функционального анализа, а в последнее время — алгоритмической и трансформационной техникой обработки текстов. Создание учения о фонологических различительных признаках, исследование морфологических процессов и типов морфологических структур, открытие свойства проективности предложений и гипотеза глубины предложения, исследование трансформационной структуры языка, открытие и исследование компонентной структуры значений обогатили наши представления о свойствах языка как семиотической системы. Наконец, исследование и систематизация языковых универсалий, а также разработка искусственных языков-эталонов послужили фундаментом для научной типологии, являющейся детищем нашего времени.

Традиционная лингвистическая проблематика была не только обогащена и углублена, но и существенным образом расширена. Действительно, в последние годы в лингвистике возникли по крайней мере две новые области. Во-первых, стали разрабатываться действующие модели языка — порождающие (имитирующие способность человека отличать правильное от неправильного в языке), аналитические и синтетические (имитирующие способность человека переходить от заданного текста к его смыслу и строить тексты по заданному смыслу). Благодаря этому лингвистика получила возможность проверять правильность своих теорий в экспериментах (поставленных на электронных вычислительных машинах), то есть применять ту методологию поиска научной истины, которая издавна используется естественными науками. Во-вторых, была создана теория грамматик, или метатеория лингвистики, мощный аппарат которой позволяет с большой точностью и надежностью оценивать и сравнивать друг с другом различные классы грамматик.

Поворот к исследованиям нового типа отнюдь не означал разрыва преемственных связей с предшествующей лингвистикой. Наоборот, новейшие исследования обещают вернуть лингвистике единство, которого она была в течение десятилетий лишена. Можно наметить несколько ярких линий преемственности, идущих от очень разных ученых,

но сходящихся в современной лингвистике: 1) от И. А. Бо-дуэна де Куртенэ и Ф. де Соссюра, разработавших основополагающие принципы синхронного описания языков, от классической описательной грамматики (Ф. Ф. Фортунатова, А. М. Пешковского, В. А. Богородицкого, Ф. Брюно, О. Есперсена, Э. Сэпира, Ш. Балли и многих других) — к Пражскому лингвистическому кружку, а затем Р. О. Якобсону и современным аналитическим, синтетическим и порождающим моделям речевой деятельности; 2) от Ф. Боаса к Л. Блумфильду, а затем американской дескриптивной лингвистике и современным исследовательским моделям; 3) от Ф. де Соссюра и его семиотических идей к Л. Ельмслеву, а затем Н. Хомскому и другим ученым, которые своими изысканиями в области метамodelей и метатеории вводят лингвистику в более широкий круг научных дисциплин, имеющих дело с разного рода искусственными языками и другими семиотическими системами.

На этом можно было бы и закончить книгу, но в заключение целесообразно порекомендовать читателю литературу, где он мог бы найти подробные сведения об интересующем его предмете. Некоторые рекомендации по частным вопросам читатель найдет в библиографических ссылках, но общее знакомство с наукой, которой посвящена эта работа, следует начать с четырех книг:

А. Т а р с к и й, Введение в логику и методологию дедуктивных наук, Изд. иностр. лит., М., 1948;

Ф. д е С о с с ю р, Курс общей лингвистики, русский перевод, Соцэкгиз, 1933;

Н. С. Т р у б е ц к о й, Основы фонологии, русский перевод, Изд. иностр. лит., М., 1960;

Н. Х о м с к и й, Синтаксические структуры, «Новое в лингвистике», вып. 2, Изд. иностр. лит., М., 1962, стр. 412—527.

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. Р. И. А в а н е с о в, Второстепенные члены предложения как грамматическая категория, «Русский язык в школе», 1936, № 1.

2. Р. И. А в а н е с о в, Фонетика современного русского литературного языка, изд. МГУ, 1956.

3. Н. Д. А н д р е е в, Моделирование языка на базе его статистической и теоретико-множественной структуры, «Тезисы совещания по математической лингвистике», Л., 1959.

4. Н. Д. А н д р е е в, Алгоритмы статистико-комбинаторного моделирования морфологии, синтаксиса, словообразования и семантики, «Материалы по математической лингвистике и машинному переводу», сб. II, Л., 1963.

5. Н. Д. А н д р е е в, Вяч. Вс. И в а н о в, И. А. М е л ь ч у к, Некоторые замечания и предложения относительно работы по машинному переводу в СССР, «Машинный перевод—и прикладная лингвистика», 1960, № 4.

6. Ю. Д. А п р е с я н, Современные методы изучения значений и некоторые проблемы структурной лингвистики, «Проблемы структурной лингвистики», изд. АН СССР, М., 1963.

7. Ю. Д. А п р е с я н, О сильном и слабом управлении (опыт количественного анализа), «Вопросы языкознания», 1964, № 3.

8. Ю. Д. А п р е с я н, Опыт описания значений глаголов по их синтаксическим признакам (типам управления), «Вопросы языкознания», 1965, № 5.

9. Ю. Д. А п р е с я н, Метод непосредственно составляющих и трансформационный метод в современной структурной лингвистике, «Русский язык в национальной школе», 1962, № 4.

10. Н. Г. А р с е н ь е в а, О двух способах порождения предложений русского языка, «Проблемы кибернетики», изд. «Наука», вып. 14, 1965.

11. Н. Г. А р с е н ь е в а, О синтезе предложений русского языка при помощи машины, «Научно-техническая информация», 1963, № 6.

12. О. С. А х м а н о в а, И. А. М е л ь ч у к, Е. В. П а д у ч е в а, Р. М. Ф р у м к и н а, О точных методах исследования языка (О так называемой «математической лингвистике»), изд. МГУ, 1961.

13. К. И. Бабицкий, Алгоритм расстановки слов во фразе при независимом русском синтезе, сб. «Машинный перевод. Труды Института ТМ и ВТ АН СССР», вып. 2, М., 1961.
14. К. И. Бабицкий, К вопросу о моделировании структуры простого предложения, «Проблемы структурной лингвистики», изд. АН СССР, М., 1962.
15. К. И. Бабицкий и Е. Л. Гинзбург, Об одном способе задания форм глагола (в печати).
16. К. И. Бабицкий, О дистрибутивной теории предложений с сочиненными частями (в печати).
17. Ш. Балли, Общая лингвистика и вопросы французского языка, Изд. иностр. лит., М., 1955.
18. И. Бар-Хиллел, О рекурсивных определениях в эмпирических науках, сб. «Математическая лингвистика», М., 1964.
19. С. И. Бернштейн, Основные понятия фонологии, «Вопросы языкознания», 1962, № 5.
20. В. А. Богородицкий, Лекции по общему языковедению, 1911.
21. В. А. Богородицкий, Общий курс русской грамматики, изд. 5, Соцэкгиз, М.—Л., 1935.
22. В. А. Богородицкий, Казанская лингвистическая школа, «Труды Московского института истории, философии и литературы», т. 5, М., 1939.
23. И. А. Бодуэн де Куртенэ, Введение в языковедение, литографированный курс 1913—1914 годов.
24. И. А. Бодуэн де Куртенэ, Сборник статей, изд. АН СССР, М., 1960.
25. И. А. Бодуэн де Куртенэ, Избранные труды по общему языкознанию, т. I, изд. АН СССР, М., 1963, т. II, изд. АН СССР, М., 1963.
26. В. И. Борковский, П. С. Кузнецов, Историческая грамматика русского языка, изд. АН СССР, 1963.
27. В. Брёндаль, Структурная лингвистика, в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 94—100.
28. Т. В. Булыгина, Пражская лингвистическая школа, в кн.: «Основные направления структурализма», изд. «Наука», М., 1964.
29. Ж. Вандриес, Язык. Лингвистическое введение в историю, Соцэкгиз, М., 1937.
30. В. В. Виноградов, Русский язык. Грамматическое учение о слове, Учпедгиз, М., 1947.
31. В. В. Виноградов, Вопросы синтаксиса русского языка в трудах М. В. Ломоносова по грамматике и риторике, «Русский язык в школе», 1950, № 2.
32. В. В. Виноградов, Синтаксические воззрения Востокова и их значение в истории русского языкознания, «Известия Академии наук. Отделение литературы и языка», т. X, № 2, 1951.
33. В. В. Виноградов, Основные типы лексических значений слова, «Вопросы языкознания», 1953, № 5.
34. В. В. Виноградов, Учение акад. А. А. Шахматова

- о грамматических формах слов и о частях речи в современном русском языке, см. сб. «Из трудов А. А. Шахматова по современному русскому языку», Учпедгиз, М., 1952.
35. В. В. Виноградов, Некоторые вопросы изучения синтаксиса простого предложения, «Вопросы языкознания», 1954, № 1.
36. Г. О. Винокур, Глагол или имя? «Русская речь», Новая серия, вып. III, Л., 1928.
37. Г. О. Винокур, Заметки по русскому словообразованию, «Избранные работы по русскому языку», Учпедгиз, М., 1959.
38. Л. Витгенштейн, Логико-философский трактат, Изд. иностр. лит., М., 1958.
39. З. М. Волоцкая, Установление отношений производности между словами, «Вопросы языкознания», 1960, № 3.
40. З. М. Волоцкая, Т. Н. Молошная, Т. М. Николеева, Опыт описания русского языка в его письменной форме, изд. «Наука», М., 1964.
41. «Вопросы глагольного вида», Изд. иностр. лит., М., 1962.
42. А. Гардинер, Различие между «речью» и «языком», в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 14—21.
43. А. Н. Гвоздев, Вопросы изучения детской речи, изд. АПН РСФСР, М., 1961.
- 43а. А. В. Гладкий, Об одном способе формализации понятия синтаксической связи, Проблемы кибернетики, изд. «Наука», М., 1964, вып. 11.
44. Г. Глисон, Введение в дескриптивную лингвистику, Изд. иностр. лит., М., 1959.
45. «Грамматика русского языка», изд. АН СССР, М., 1960.
46. Э. Добльхофер, Знаки и чудеса, Изд. восточной литературы, М., 1963.
47. Р. Л. Добрушин, Элементарная грамматическая категория, «Бюллетень объединения по проблемам машинного перевода», вып. 5, 1957.
48. Р. Л. Добрушин, Математические методы в лингвистике, «Математическое просвещение», вып. 6, М., 1961.
49. А. Б. Долгопольский, Гипотеза древнейшего родства языковых семей северной Евразии с вероятностной точки зрения, «Вопросы языкознания», 1964, № 2.
50. Ф. А. Дрейзин, Об одном способе синтаксического анализа простого предложения, «Научные труды Ташкентского государственного университета», вып. 208, 1962.
51. Л. Ельмслев, Метод структурного анализа в лингвистике, в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 103—110.
52. Л. Ельмслев, Язык и речь, в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 111—120.
53. Л. Ельмслев, Прологомены к теории языка, «Новое в лингвистике», т. I, Изд. иностр. лит., М., 1960.
54. Л. Ельмслев, Можно ли считать, что значения слов

образуют структуру? «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

55. О. Есперсен. Философия грамматики, Изд. иностр. лит., М., 1958.

56. А. К. Жолковский, Работы Э. Сэпира по структурной семантике, «Машинный перевод и прикладная лингвистика», 1964, № 8.

57. А. К. Жолковский, О правилах семантического анализа, «Машинный перевод и прикладная лингвистика», 1964, № 8.

58. А. К. Жолковский, Лексика целесообразной деятельности, «Машинный перевод и прикладная лингвистика», 1964, № 8.

59. А. К. Жолковский, Н. Н. Леонтьева, Ю. С. Мартемьянов, О принципиальном использовании смысла при машинном переводе, «Машинный перевод. Труды Института ТМ и ВТ АН СССР», вып. 2, М., 1961.

60. А. А. Зализняк, О возможной связи между операционными понятиями синхронного описания и диахронией, «Симпозиум по структурному изучению знаковых систем», М., 1962.

61. А. А. Зализняк, Лингвистические задачи, сб. «Исследования по структурной типологии», изд. АН СССР, М., 1963.

62. А. А. Зализняк, К вопросу о грамматической категории рода и одушевленности в современном русском языке, «Вопросы языкознания», 1964, № 4.

62а. А. А. Зализняк, Классификация и синтез именных парадигм современного русского языка, Диссертация на соискание ученой степени доктора филологических наук, М., 1965.

63. Л. Н. Засорина, Трансформации как метод лингвистического эксперимента в синтаксисе, сб. «Трансформационный метод в структурной лингвистике», изд. «Наука», М., 1964.

64. Л. Н. Засорина, «Порядок слов» при синтезе русского предложения, «Материалы по математической лингвистике и машинному переводу», вып. II, Л., 1963.

65. В. А. Звегинцев, Дескриптивная лингвистика, предисловие к кн. Г. Глисона (см. № 44).

66. В. А. Звегинцев, Глоссематика и лингвистика, сб. «Новое в лингвистике», т. I, Изд. иностр. лит., М., 1960.

67. Е. А. Земская, Об одной особенности соединения словообразовательных морфем в русском языке, «Вопросы языкознания», 1964, № 2.

68. Г. А. Золотова, Развитие некоторых типов именных двусоставных предложений в современном русском языке, «Развитие грамматики и лексики современного русского языка», изд. «Наука», М., 1964.

69. В. ч. В. С. Иванов, Некоторые соображения о трансформационной грамматике, «Тезисы докладов на конференции по структурной лингвистике», М., 1961.

70. В. ч. В. С. Иванов, Теория фонологических различительных признаков, «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

71. В. ч. В. С. Иванов, Лингвистика и исследование афазии, сб. «Структурно-типологические исследования», изд. АН СССР, М., 1962.

72. Л. И. И л и я, Синтаксис современного французского языка, М., 1962.
73. В. И н г в е, Синтаксис и проблема многозначности, сб. «Машинный перевод», М., 1957.
74. Л. Н. И о р д а н с к а я, Два оператора для обработки словосочетаний с «сильным управлением» (для автоматического синтаксического анализа), М., 1961.
75. Л. Н. И о р д а н с к а я, О некоторых свойствах правильной синтаксической структуры, «Вопросы языкознания», 1963, № 4.
76. Л. Н. И о р д а н с к а я, Свойства правильной синтаксической структуры и алгоритм ее обнаружения, «Проблемы кибернетики», вып. II, М., 1964.
77. А. Ф. И о ф ф е, Основные представления современной физики, М.—Л., 1949.
78. А. В. И с а ч е н к о, Грамматический строй русского языка в сопоставлении с словацким, Братислава, 1960.
79. А. В. И с а ч е н к о, О грамматическом значении, «Вопросы языкознания», 1961, № 1.
80. А. В. И с а ч е н к о, Бинарность, привативные оппозиции и грамматические значения, «Вопросы языкознания», 1963, № 2.
81. А. В. И с а ч е н к о, Трансформационный анализ кратких и полных прилагательных, «Исследования по структурной лингвистике», изд. АН СССР, М., 1963.
82. А. Н. К о л м о г о р о в, Автоматы и жизнь, сб. «Возможное и невозможное в кибернетике», М., 1964.
83. Б. Г. К у з н е ц о в, Эйнштейн, изд. АН СССР, М., 1963.
84. П. С. К у з н е ц о в, О последовательности построения системы языка, «Тезисы конференции по машинному переводу», М., 1958.
85. П. С. К у з н е ц о в, Об основных положениях фонологии, «Вопросы языкознания», 1959, № 2.
86. П. С. К у з н е ц о в, О принципах изучения грамматики, изд. МГУ, 1961.
87. П. С. К у з н е ц о в, Опыт формального определения слова, «Вопросы языкознания», 1964, № 5.
88. О. С. К у л а г и н а, Об одном способе определения грамматических понятий на базе теории множеств, сб. «Проблемы кибернетики», вып. I, М., 1958.
89. Е. К у р и л о в и ч, Деривация лексическая и деривация синтаксическая, «Очерки по лингвистике», Изд. иностр. лит., М., 1962.
90. Е. К у р и л о в и ч, Понятие изоморфизма, там же.
91. Е. К у р и л о в и ч, Основные структуры языка: словосочетание и предложение, там же.
92. Е. К у р и л о в и ч, Проблема классификации падежей, там же.
93. А. Г. К у р о ш, Лекции по общей алгебре, М., 1962.
94. К. А. Л е в к о в с к а я, О принципах структурно-семантического анализа языковых единиц, «Вопросы языкознания», 1957, № 1.
95. Ю. К. Л е к о м ц е в, Основные положения глоссематики, «Вопросы языкознания», 1962, № 4.

96. А. А. Леонтьев, Общелингвистические взгляды И. А. Бодуэна де Куртенэ (к 30-летию со дня смерти), «Вопросы языкознания», 1959, № 6.

97. Н. Н. Леонтьева, Модель синтеза русской фразы на основе семантической записи, «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текстов», вып. 9, М., 1961.

98. И. Лесерф, Применение программы и модели конфликтной ситуации к автоматическому синтаксическому анализу, «Научно-техническая информация», 1963, № 10.

99. Г. А. Лесский, О некоторых различиях простого предложения в научной и художественной прозе, «Русский язык в национальной школе», 1963, № 6.

100. Р. Б. Лиз, Что такое трансформация? «Вопросы языкознания», 1961, № 3.

101. Р. Б. Лиз, О переформулировании трансформационных грамматик, «Вопросы языкознания», 1961, № 6.

102. А. А. Ляпунов, Об управляющих системах живой природы и общем понимании жизненных процессов, сб. «Проблемы кибернетики», 1963, вып. 10.

103. Ю. С. Мартеньянов, О кодировке слов для алгоритма автоматического синтаксического анализа, «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текстов», вып. 10, М., 1961.

104. А. Мартине, Принцип экономии в фонетических изменениях, Изд. иностр. лит., М., 1960.

105. А. Мартине, О книге «Основы лингвистической теории» Л. Ельмслева, «Новое в лингвистике», т. I, Изд. иностр. лит., 1960.

106. А. Мартине, Основы общей лингвистики, «Новое в лингвистике», т. III, Изд. иностр. лит., М., 1963.

107. Ж. Марузо, Словарь лингвистических терминов, Изд. иностр. лит., М., 1960.

108. М. Мастерман, Тезаурус в синтаксисе и семантике, сб. «Математическая лингвистика», М., 1964.

109. М. Мастерман, Изучение семантической структуры текста для машинного перевода с помощью языка-посредника, сб. «Математическая лингвистика», М., 1964.

110. В. Матезиус, Куда мы пришли в языкознании, в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 141—146.

110а. «Машинный перевод и прикладная лингвистика», М., № 8, 1964.

111. И. А. Мельчук, Статистика и зависимость рода французских существительных от их окончаний, сб. «Вопросы статистики речи», Л., 1958.

112. И. А. Мельчук, О терминах «устойчивость» и «идиоматичность», «Вопросы языкознания», 1960, № 4.

113. И. А. Мельчук, К вопросу о «грамматическом» в языке-посреднике, «Машинный перевод и прикладная лингвистика», М., 1960, № 4.

114. И. А. Мельчук, Некоторые вопросы машинного пе-

ревода за рубежом, «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текстов», вып. 6, М., 1961.

115. И. А. М е л ь ч у к, К вопросу о термине «система» в лингвистике, «Zeichen und System der Sprache», Band II, Berlin, 1962.

116. И. А. М е л ь ч у к, Об алгоритме синтаксического анализа языковых текстов (общие принципы и некоторые итоги), «Машинный перевод и прикладная лингвистика», 1962, № 7.

117. И. А. М е л ь ч у к, О стандартной форме и количественных характеристиках некоторых лингвистических описаний, «Вопросы языкознания», 1963, № 1.

118. И. А. М е л ь ч у к, Автоматический анализ текстов, сб. «Славянское языкознание», Доклады советской делегации, изд. АН СССР, М., 1963.

119. И. И. М е щ а н и н о в, Члены предложения и части речи, изд. АН СССР, М.—Л., 1945.

120. Т. Н. М о л о ш н а я, Алгоритм перевода с английского языка на русский, «Проблемы кибернетики», вып. 3, 1960.

121. Т. Н. М о л о ш н а я, О понятии грамматической конфигурации, «Структурно-типологические исследования», изд. АН СССР, М., 1962.

122. Т. Н. М о л о ш н а я, Грамматические трансформации английского языка, «Трансформационный метод в структурной лингвистике», изд. АН СССР, М., 1964.

123. Т. Н. М о л о ш н а я, Вопросы различия омонимов при машинном переводе с английского языка на русский, «Проблемы кибернетики», 1958, вып. 1.

124. Т. М. Н и к о л а е в а, Построение предложения при независимом синтезе русского текста, «Машинный перевод. Труды Института ТМ и ВТ АН СССР», вып. 2, М., 1961.

125. Т. М. Н и к о л а е в а, Письменная речь и специфика ее изучения, «Вопросы языкознания», 1961, № 3.

126. Т. М. Н и к о л а е в а, Опыт алгоритмической морфологии русского языка, «Структурно-типологические исследования», изд. АН СССР, М., 1962.

127. Т. М. Н и к о л а е в а, Трансформационный анализ прилагательных с прилагательным — управляющим словом, «Трансформационный метод в структурной лингвистике», изд. АН СССР, М., 1964.

128. С. И. О ж е г о в, Словарь русского языка, изд. 4, М., 1960.

129. «Основные направления структурализма», изд. «Наука», М., 1964.

130. И. П. П а в л о в, Естествознание и мозг, «Избранные труды», М., 1950.

131. Е. В. П а д у ч е в а, Об описании падежной системы русского существительного, «Вопросы языкознания», 1960, № 5.

132. Е. В. П а д у ч е в а, Классификация сложных предложений в связи с построением правил образования для стандартизованного русского языка, «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текстов», М., 1961.

133. Е. В. Падучева, О способах представления синтаксической структуры предложения, «Вопросы языкознания», 1964, № 2.

134. Е. В. Падучева и А. Л. Шумилина, Описание синтагм русского языка (в связи с построением алгоритма машинного перевода), «Вопросы языкознания», 1961, № 4.

135. М. В. Панов, О слове как единице языка, «Ученые записки Московского городского педагогического института им. В. П. Потемкина», т. 51, вып. 5, 1956.

136. М. В. Панов, Словообразование, в кн.: «Русский язык и советское общество», Алма-Ата, 1962.

137. М. В. Панов, Синтаксис, в кн.: «Русский язык и советское общество», Алма-Ата, 1962.

138. М. В. Панов, О некоторых общих тенденциях в развитии русского литературного языка XX века (основные позиционные изменения в фонетике и морфологии), «Вопросы языкознания», 1963, № 1.

139. Ф. Пап, Трансформационный анализ русских присубстантивных конструкций с зависимой частью — существительным, «Publicationes instituti philologiae slavicae universitatis debrecenensis», Debrecen, 1961.

140. В. Пекеллис, Музы и машины, сб. «Возможное и невозможное в кибернетике», М., 1964.

141. А. М. Пешковский, Еще к вопросу о предмете синтаксиса, «Русский язык в советской школе», 1929, № 2.

142. А. М. Пешковский, Глагольность как выразительное средство, сб. «Методика родного языка, лингвистика, стилистика, поэтика», Л., 1925; см. также: А. М. Пешковский, Избранные труды, Учпедгиз, М., 1959, стр. 101—111.

143. А. М. Пешковский, Существует ли в русском языке сочинение и подчинение предложений? Сб. «Вопросы методики родного языка, лингвистики и стилистики», М.—Л., 1930; см. также: А. М. Пешковский, Избранные труды, Учпедгиз, М., 1959, стр. 131—146.

144. А. М. Пешковский, Русский синтаксис в научном освещении, изд. 5, Учпедгиз, М., 1935.

145. Е. Д. Поливанов, За марксистское языкознание, изд. «Федерация», М., 1931.

146. Л. Э. Пшеничная и Э. Ф. Скороходько, Синтез осмысленных предложений на ЭЦВМ, «Проблемы кибернетики», вып. 10, 1963.

147. Р. Д. Равич, Некоторые данные о состоянии машинного перевода, «Научно-техническая информация», 1964, № 5—6.

148. И. И. Ревзин, О соотношении структурных и статистических методов в современной лингвистике, сб. «Вопросы статистики речи», Л., 1958.

149. И. И. Ревзин, «Активная» и «пассивная» грамматика Л. В. Щербы и проблемы машинного перевода, «Тезисы конференции по машинному переводу», М., 1958.

150. И. И. Ревзин, О сильных и слабых противопоставлениях в системе падежей современного немецкого языка, «Вопросы языкознания», 1960, № 3.

151. И. И. Ревзин, О некоторых понятиях теоретико-

множественной концепции языка, «Вопросы языкознания», 1960, № 6.

152. И. И. Ревзин, Модели языка, изд. АН СССР, М., 1962.

153. И. И. Ревзин, О некоторых вопросах дистрибутивного анализа и его дальнейшей формализации, «Проблемы структурной лингвистики», изд. АН СССР, М., 1962.

154. И. И. Ревзин и В. Ю. Розенцвейг, Основы общего и машинного перевода, М., 1964.

155. А. А. Реформатский, Введение в языкознание, изд. 2, Учпедгиз, М., 1955.

156. А. А. Реформатский, Что такое структурализм? «Вопросы языкознания», 1957, № 6.

157. А. А. Реформатский, Н. С. Трубецкой и его «Основы фонологии», в кн.: Н. С. Трубецкой, Основы фонологии, Изд. иностр. лит., М., 1960.

158. А. А. Реформатский, Дихотомическая классификация дифференциальных признаков и фонематическая модель языка, сб. «Вопросы теории языка в современной зарубежной лингвистике», изд. АН СССР, М., 1961.

158а. Б. С. Роговой, Лингвopsихологические эксперименты с деграмматизованными текстами, сб. «Вопросы общего языкознания», Л., 1965.

159. Р. Ружичка, О трансформационном описании так называемых безличных предложений в современном русском литературном языке, «Вопросы языкознания», 1963, № 3.

160. М. В. Сергиевский, Современные грамматические теории в Западной Европе и античная грамматика, «Ученые записки И МГПИИЯ», т. II, 1940.

161. В. Н. Сидоров и И. С. Ильинская, К вопросу о выражении субъекта и объекта действия в современном русском литературном языке, «Известия Академии наук СССР. Отделение литературы и языка», т. VIII, вып. 4, 1949.

162. В. Скаличка, Копенгагенский структурализм и «Пражская школа», в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 147—154.

163а. Н. Д. Слюсарева, Лингвистический анализ по непосредственно составляющим, «Вопросы языкознания», 1960, № 6.

164. Ф. де Соссюр, Курс общей лингвистики, М., 1933.

165. А. И. Смирницкий, Некоторые замечания о принципах морфологического анализа основ, «Доклады и сообщения филологического факультета МГУ», вып. 5, 1948.

166. А. И. Смирницкий, К вопросу о слове (проблема тождества слова), «Труды института языкознания», т. IV, изд. АН СССР, М., 1954.

167. А. И. Смирницкий, Лексикология английского языка, М., 1956.

168. А. И. Смирницкий, Синтаксис английского языка, М., 1957.

169. А. И. Смирницкий, Морфология английского языка, М., 1959.

170. П. А. Соболева, О трансформационном анализе

словообразовательных отношений, сб. «Трансформационный метод в структурной лингвистике», изд. «Наука», М., 1964.

171. «Сравнительная грамматика германских языков», т. II, изд. АН СССР, М., 1962.

172. М. И. Стеблин-Каменский, Несколько замечаний о структурализме, «Вопросы языкознания», 1957, № 1.

173. Б. В. Сухотин, Экспериментальное выделение классов букв с помощью электронной вычислительной машины, «Проблемы структурной лингвистики», изд. АН СССР, М., 1962.

174. Б. В. Сухотин, Алгоритмы лингвистической дешифровки, «Проблемы структурной лингвистики», изд. АН СССР, М., 1963.

175. Э. Сэпир, Язык. Введение в изучение речи, Соцэкгиз, М., 1934.

176. «Тезисы Пражского лингвистического кружка», в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 123—140.

177. В. Н. Топоров, О введении вероятности в языкознание, «Вопросы языкознания», 1960, № 6.

178. Б. Трнка, К дискуссии по вопросам структурализма, в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 155—166.

179. Н. С. Трубецкой, Основы фонологии, Изд. иностр. лит., М., 1960.

180. Д. С. Уорс, Трансформационный анализ конструкций с творительным падежом в русском языке, сб. «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

181. Д. С. Уорс, Об отображении линейных отношений в порождающих моделях языка, «Вопросы языкознания», 1964, № 5.

182. Б. Л. Уорф, Отношение норм поведения и мышления к языку, сб. «Новое в лингвистике», т. I, Изд. иностр. лит., М., 1960.

183. Б. Л. Уорф, Наука и языкознание, сб. «Новое в лингвистике», т. I, Изд. иностр. лит., М., 1960.

184. Б. Л. Уорф, Лингвистика и логика, сб. «Новое в лингвистике», т. I, Изд. иностр. лит., М., 1960.

185. Б. А. Успенский, Рец. на кн. «Universals of Language», «Вопросы языкознания», 1963, № 5.

186. В. А. Успенский, К определению части речи в теоретико-множественной системе языка, «Бюллетень объединения по проблемам машинного перевода», вып. 5, 1957.

187. Л. Успенский, Слово о словах, Детгиз, М., 1957.

188. Г. К. Уэллс, Павлов и Фрейд, Изд. иностр. лит., М., 1959.

189. С. Я. Фиталов, О моделировании синтаксиса в структурной лингвистике, «Проблемы структурной лингвистики», изд. АН СССР, М., 1962.

190. С. Я. Фиталов, Трансформации в аксиоматических грамматиках, «Трансформационный метод в структурной лингвистике», изд. «Наука», М., 1964.

191. Ф. Ф. Фортунатов, Сравнительное языковедение

(общий курс лекций, читанных в 1901—1902 акад. году, литогр. издание); см. также: Ф. Ф. Фортунатов, Избранные труды, т. I, Учпедгиз, М., 1956, стр. 23—199.

192. Р. М. Фрумкина, Автоматизация исследовательских работ в области лексикологии и лексикографии, «Вопросы языкознания», 1964, № 2.

193. Р. М. Фрумкина, Статистические методы изучения лексики, М., 1964.

194. М. Халле, О правилах русского спряжения, «American Contributions to the Fifth International Congress of Slavists», The Hague, 1963.

195. З. С. Харрис, Совместная встречаемость и трансформация в языковой структуре, «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

196. Э. Хауген, Направления в современном языкознании, «Новое в лингвистике», т. I, Изд. иностр. лит., М., 1960.

197. А. Хилл, О грамматической отмеченности предложений, «Вопросы языкознания», 1962, № 4.

198. А. А. Холодович, Опыт теории подклассов слов, «Вопросы языкознания», 1960, № 1.

199. Н. Хомский, Синтаксические структуры, «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

200. Н. Хомский, Лингвистика, логика, психология и вычислительные устройства, «Математическая лингвистика», М., 1964.

201. Н. Хомский, Несколько методологических замечаний о порождающей грамматике, «Вопросы языкознания», 1962, № 4.

202. Н. Хомский, Три модели описания языка, «Кибернетический сборник», вып. 2, М., 1963.

203. А. Я. Шайкевич, Распределение слов в тексте и выделение семантических полей, «Иностранные языки в высшей школе», 1963.

204. С. К. Шаумян, Теоретические основы трансформационной грамматики, «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

205. С. К. Шаумян, Проблемы теоретической фонологии, изд. АН СССР, М., 1962.

206. С. К. Шаумян, Насущные задачи структурной лингвистики, «Известия Академии наук СССР. Отделение литературы и языка», т. XXI, вып. 2, 1962.

207. С. К. Шаумян, Преобразование информации в процессе познания и двухступенчатая теория структурной лингвистики, «Проблемы структурной лингвистики», изд. АН СССР, М., 1962.

208. С. К. Шаумян, Порождающая лингвистическая модель на базе принципа двухступенчатости, «Вопросы языкознания», 1963, № 2.

209. С. К. Шаумян, Трансформационная грамматика и аппликативная порождающая модель, «Трансформационный метод в структурной лингвистике», изд. «Наука», М., 1964.

210. С. К. Шаумян, Основы структурной лингвистики, изд. «Наука», М., 1965.

211. С. К. Шаумян и П. А. Соболева, Аппликативная порождающая модель и исчисление трансформаций в русском языке, изд. АН СССР, М., 1963.

212. А. А. Шахматов, Синтаксис русского языка, вып. 2, изд. АН СССР, 1927.

213. А. А. Шахматов, Синтаксис русского языка, Учпедгиз, М.—Л., 1941.

214. Н. Ю. Шведова, Некоторые виды значений сказуемого в современном русском языке, сб. «Исследования по грамматике русского литературного языка», изд. АН СССР, М., 1955.

215. Н. Ю. Шведова, Проблема лексических ограничений как одна из проблем изучения истории синтаксиса русского литературного языка XVIII—XIX веков, «Вопросы языкознания», 1960, № 6.

216. Н. Ю. Шведова, О некоторых активных процессах в современном русском синтаксисе (наблюдения над языком газеты), «Вопросы языкознания», 1964, № 2.

217. Н. Ю. Шведова, Парадигматика простого предложения в современном русском языке (опыт типологии), сб. «Русский язык. Грамматические исследования» (в печати).

218. В. В. Шеворошкин, Карийский вопрос, «Вопросы языкознания», 1962, № 5.

219. В. В. Шеворошкин, О структуре звуковых цепей, «Проблемы структурной лингвистики», изд. АН СССР, М., 1963.

220. Д. Н. Шмелев, О «связанных» синтаксических конструкциях в русском языке, «Вопросы языкознания», 1960, № 5.

221. Ю. А. Шрейдер, Машинный перевод на основе смыслового кодирования текстов, «Научно-техническая информация», 1963, № 1.

222. Э. А. Штейнфельдт, Частотный словарь современного русского литературного языка, Таллин, 1963.

223. Л. В. Щерба, И. А. Бодуэн де Куртенэ и его значение в науке о языке, «Избранные работы по русскому языку», Учпедгиз, М., 1957.

224. Л. В. Щерба, О частях речи в русском языке, «Избранные работы по русскому языку», Учпедгиз, М., 1957.

225. Л. В. Щерба, О тройном аспекте языковых явлений и об эксперименте в языкознании, в кн.: В. А. Звегинцев, История языкознания XIX—XX веков в очерках и извлечениях, ч. II, изд. 3, изд. «Просвещение», М., 1965, стр. 361—372.

226. Л. В. Щерба, О второстепенных членах предложения, «Избранные работы по языкознанию и фонетике», изд. ЛГУ, т. I, 1958.

227. Л. В. Щерба, Очередные проблемы языковедения, «Избранные работы по языкознанию и фонетике», изд. ЛГУ, т. I, 1958.

228. Л. В. Щерба, Ф. Ф. Фортунатов в истории науки о языке, «Вопросы языкознания», 1963, № 5.

229. Л. В. Щерба, Преподавание иностранных языков в средней школе, М., 1947.

230. Л. В. Щерба, Фонетика французского языка. Очерк французского произношения в сравнении с русским, М., 1948.

231. А. Эйштейн и Л. Инфельд, Эволюция физики, М.—Л., 1948.

232. У. Р. Эшби, Введение в кибернетику, М., 1959.

233. У. Р. Эшби, Что такое разумная машина, сб. «Возможное и невозможное в кибернетике», М., 1964.

234. И. М. Яглом, Р. Л. Добрушин, А. М. Яглом, Теория информации и лингвистика, «Вопросы языкознания», 1960, № 1.

235. Р. Якобсон, Г. М. Фанти и М. Халле, Введение в анализ речи, «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

236. Р. Якобсон, М. Халле, Фонология и ее отношение к фонетике, «Новое в лингвистике», т. II, Изд. иностр. лит., М., 1962.

237. Н. А. Янко-Триницкая, Возвратные глаголы в современном русском языке, изд. АН СССР, М., 1962.

238. *Anthropology Today, An Encyclopedic Inventory Prepared under the Chairmanship of A. L. Kroeber*, Chicago, 1957.

239. E. Bach, Subcategories in Transformational Grammars, «Preprints of Papers for the Ninth International Congress of Linguists», Cambridge, Mass., 1962.

240. E. Bach, An Introduction to Transformational Grammars, N. Y., Chicago, San-Francisco, 1964.

241. Ch. Bally, «La pensée et la langue» de F. Brunot, «Bulletin de la Société de linguistique de Paris», t. 23, 3 fasc., Paris, 1922.

242. Ch. Bally, *Traité de stylistique française*, Paris, 2-nde éd.

243. Y. Bar-Hillel, A Quasi-arithmetical Notation for Syntactic Description, «Language», v. 29, No. 1, 1953.

244. Y. Bar-Hillel, Logical Syntax and Semantics, «Language», v. 30, No. 2, 1954.

245. Y. Bar-Hillel, Report on the State of MT in the USA and Great Britain, Jerusalem, 1959.

246. E. Benveniste, Problèmes sémantiques de la reconstruction, «Word», v. 10, № 2—3, 1954.

247. E. Benveniste, Les niveaux de l'analyse linguistique, «Preprints of Papers for the Ninth International Congress of Linguists», Cambridge, Mass., 1962.

248. B. Bloch, English Verb Inflection, «Readings in Linguistics», edited by M. Joos, Washington, 1957.

249. B. Bloch, A Set of Postulates for Phonemic Analysis, «Language», v. 24, No. 1, 1948.

250. B. Bloch, Studies in Colloquial Japanese, «Language», v. 22, No. 2, 1946.

251. B. Bloch and G. L. Trager, Outline of Linguistic Analysis, Baltimore, 1942.

252. L. Bloomfield, A Set of Postulates for the Science of Language, «Readings in Linguistics», edited by M. Joos, Washington, 1957.

253. L. Bloomfield, Language, N. Y., 1933.

253a. L. Bloomfield, Meaning, «Monatshefte für deutschen Unterricht», b. 35, 1943.

254. F. Boas, Handbook of American Indian Languages, v. I, Washington, 1911; v. II, Washington, 1922.

255. D. Bolinger, Syntactic Blends and Other Matters, «Language», v. 37, No. 3, 1961.
256. V. Brøndal, Essais de linguistique générale, Copenhagen, 1943.
257. F. Brunot, La pensée et la langue, Paris, 1922.
258. F. Brunot et Ch. Bruneau, Précis de grammaire historique de la langue française, Paris, 1933.
259. J. B. Carroll, The Study of Language, A Survey of Linguistics and Related Disciplines in America, Cambridge, Mass., 1953.
260. S. Chatman, Immediate Constituents and Expansion Analysis, «Word», v. II, No. 3, 1955.
261. N. Chomsky, Logical Syntax and Semantics. Their Linguistic Relevance, «Language», v. 31, No. 1, 1955.
262. N. Chomsky, On the Notion «Rule of Grammar», «Proceedings of Symposia of Applied Mathematics», v. 12.
263. N. Chomsky, Logical Structure in Language, «American Documentation», v. VIII, No. 4, 1957.
264. N. Chomsky, The Logical Basis of Linguistic Theory, «Preprints of Papers for the Ninth International Congress of Linguists», Cambridge, Mass., 1962.
265. N. Chomsky, G. A. Miller, Introduction to the Formal Analysis of Natural Languages, «Handbook of Mathematical Psychology», v. 2, N. Y., 1963.
266. Contemporary Approaches to Cognition, Cambridge, Mass., 1957.
267. F. Daneš — J. Vachek, Prague Studies in Structural Grammar Today, «Travaux linguistiques de Prague», Prague, 1964.
268. A. Dauzat, Grammaire raisonnée de la langue française, Lyon, 1947.
269. P. Diderichsen, The Importance of Distribution versus Other Criteria in Linguistic Analysis, «Reports for the Eighth International Congress of Linguists», v. I, Oslo, 1957.
270. A. R. Diebold Jr., Psycholinguistics: A Book of Readings, ed. by Sol Saporta, N. Y., 1961.
271. C. L. Ebeling, Linguistic Units, 's-Gravenhage, 1960.
272. H. P. Edmunsden, A Statistician's View of Linguistic Models and Language—data Processing, «Natural Language and the Computer», N. Y., 1963.
273. A. Frei, La grammaire des fautes, Paris, 1929.
274. Ch. Fries, The Structure of English, N. Y., 1952.
275. A. Gardiner, The Theory of Speech and Language, London, 1951.
276. P. L. Garvin, The Definitional Model of Language, «Natural Language and the Computer», N. Y., 1963.
277. P. L. Garvin, Syntax in Machine Translation, «Natural Language and the Computer», 1963, N. Y.
278. P. L. Garvin, On Linguistic Method. Selected Papers The Hague, 1964.
279. P. L. Garvin and W. Karush, Linguistics, Data Processing, and mathematics, «Natural Language and the Computer», N. Y., 1963.

280. J. H. Greenberg, *Essays in Linguistics*, Chicago, 1957.
281. R. Godel, *Les sources manuscrites du «Cours de linguistique générale» de F. de-Saussure*, Genève — Paris, 1957.
282. A. W. de Groot, *Les oppositions dans les systèmes de la syntaxe et des cas*, «*Melanges Bally*», 1939.
283. A. W. de Groot, *Structural Linguistics and Word Classes*, «*Lingua*», v. 1, No. 1, 1948.
284. A. W. de Groot, *Structural Linguistics and Syntactic Laws*, «*Word*», v. 5, No. 1, 1949.
285. W. Haas, *Linguistic Structures*, «*Word*», v. 16, No. 2, 1960.
286. M. Halle, *On the Role of Simplicity in Linguistic Descriptions*, «*Structure of Language and its Mathematical Aspects*», N. Y., 1961.
287. M. Halliday, *Categories of the Theory of Grammar*, «*Word*», v. 17, No. 3, 1961.
288. Z. S. Harris, *Morpheme Alternants in Linguistic Analysis*, «*Readings in Linguistics*», edited by M. Joos, Washington, 1957.
289. Z. S. Harris, *From Morpheme to Utterance*, «*Readings in Linguistics*», edited by M. Joos, Washington, 1957.
290. Z. S. Harris, *Discontinuous Morphemes*, «*Language*», v. 21, No. 3, 1945.
291. Z. S. Harris, *Componential Analysis of a Hebrew Paradigm*, «*Language*», v. 24, No. 1, 1948.
292. Z. S. Harris, *Discourse Analysis*, «*Language*», v. 28, No. 1, 1952.
293. Z. S. Harris, *Distributional Structure*, «*Word*», v. 10, No. 2—3, 1954.
294. Z. S. Harris, *From Phoneme to Morpheme*, «*Language*», v. 31, No. 2, 1955.
295. Z. S. Harris, *Structural Linguistics*, Chicago, 1961.
296. Z. S. Harris, *String Analysis of Sentence Structure*, The Hague, 1962.
297. F. W. Harwood, *Axiomatic Syntax. The Construction and Evaluation of a Syntactic Calculus*, «*Language*», v. 31, No. 3, 1955.
298. R. M. Hayes, *Research Procedures in Machine Translation*, «*Natural Language and the Computer*», N. Y., 1963.
299. R. M. Hayes, *Mathematical Models for Information Retrieval*, «*Natural Language and the Computer*», N. Y., 1963.
300. A. Hill, *Introduction to Linguistic Structures*, N. Y., 1958.
301. A. Hill, *Suprasegmentals, Prosodies, Prosodemes*, «*Language*», v. 37, No. 4, 1961.
302. L. Hjelmslev, *La catégorie des cas*, Aarhus, 1936.
303. L. Hjelmslev, *La stratification du langage*, «*Word*», v. 10, No. 2—3, 1954.
304. L. Hjelmslev, H. J. Uldall, *Outline of Glossematics. A Study in the Methodology of the Humanities with Special Reference to Linguistics*, «*Travaux du cercle linguistique de Copenhague*», v. X₁, Copenhagen, 1957.
305. Ch. Hockett, *Problems of Morphemic Analysis*, «*Readings in Linguistics*», edited by M. Joos, Washington, 1957.

306. Ch. H o c k e t t, Two Fundamental Problems in Phonemics, «Readings in Linguistics», edited by M. Joos, Washington, 1957.
307. Ch. H o c k e t t, Two Models of Grammatical Description, «Word», v. 10, No. 2—3, 1954.
308. Ch. H o c k e t t, A Manual of Phonology, Baltimore, 1955.
309. Ch. H o c k e t t, A Course in Modern Linguistics, N. Y., 1958.
310. Ch. H o c k e t t, Linguistic Elements and Their Relations, «Language», v. 37, No. 1, 1961.
311. Ch. H o c k e t t, Grammar for the Hearer, «Structure of Language and its Mathematical Aspects», N. Y., 1961.
312. H. H o i j e r, Cultural Implications of Some Navaho Linguistic Categories, «Language», v. 27, No. 2, 1951.
313. H. H o i j e r, The Sapir-Whorf Hypothesis, «Language in Culture», Chicago, 1954.
314. H. H o i j e r, The Relation of Language to Culture, «Anthropology Today. An Encyclopedic Inventory Prepared under the Chairmanship of A. L. Kroeber», Chicago, 1957.
315. A. V. I s a č e n k o und H.-J. S c h ä d l i c h, Erzeugung künstlicher deutscher Satzintonationen mit zwei kontrastierenden Tonstufen, «Monatsberichte der Deutschen Akademie der Wissenschaften zu Berlin», Band 5, Heft 6, 1963.
316. R. J a k o b s o n, Zur Struktur des russischen Verbums, «Charisteria Guilermo Mathesio quinquagenario a discipulis et circuli linguistici pragensis sodalibus oblata», Pragae, 1932.
317. R. J a k o b s o n, Beitrag zur Allgemeine Kasuslehre, «Travaux du cercle linguistique de Prague», v. 6, 1936.
318. R. J a k o b s o n, Linguistics and Communication. Theory, «Structure of Language and its Mathematical Aspects», N. Y., 1961.
319. O. J e s p e r s e n, A Modern English Grammar on Historical Principles, Heidelberg, 1927, v. V, Syntax.
320. O. J e s p e r s e n, Essentials of English Grammar, 1933.
321. D. J o n e s, Outline of English Phonetics, London, 1936.
322. H. J o s s e l s o n, The Russian Word Count, Detroit, 1953.
323. A. J u i l l a n d, Outline of a General Theory of Structural Relations, «Janua linguarum», 's-Gravenhage, 1961.
324. S. K a r c e v s k i, Système du verbe russe, Prague, 1927.
325. J. J. K a t z, Mentalism in Linguistics, «Language», v. 40, No. 2, 1964.
- 325a. T. T. K a t z, P. M. P o s t a l, An Integrated Theory of Linguistic Descriptions, Cambridge (Mass.), 1964.
326. Ch. K l u c k h o h n, Notes on Some Anthropological Aspects of Communication, «American Anthropologist», v. 63, No. 5, 1961.
327. «Language in Culture», Chicago, 1954.
328. R. L e e s, рец. на кн.: N. C h o m s k y, «Syntactic Structures», «Language», v. 33, 1957.
329. R. L e e s, The Grammar of English Nominalizations, «International Journal of American Linguistics», v. XXVI, No. 3, p. II, 1960.

330. R. Lees, A Multiply Ambiguous Adjectival Construction in English, «Language», v. 36, No. 2, 1960.
331. E. Lenneberg, Cognition in Ethnolinguistics, «Language», v. 29, No. 4, 1953.
332. R. Longacre, String Constituent Analysis, «Language», v. 36, No. 1, 1960.
333. S. Marcus, Lingvistică matematică. Modele matematice în lingvistică, București, 1963.
334. A. Martiné, The Unity of Linguistics, «Word», v. 10, No. 2—3, 1954.
335. A. Martiné, Structural Linguistics, «Anthropology Today. An Encyclopedic Inventory Prepared under the Chairmanship of A. L. Croeber», Chicago, 1957.
336. M. Masterman, Semantic Problems in Language, Colloquium Report, Cambridge, England, 1962.
337. V. Mathesius, On Some Problems of the Systematic Analysis of Grammar, «Travaux du cercle linguistique de Prague», v. 6, 1936.
338. J. Mersel, Programming Aspects of Machine Translation, «Natural Language and the Computer», N. Y., 1963.
339. G. A. Miller, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, «Psychological Review», v. 63, 1956.
340. G. A. Miller and N. Chomsky, Finitary Models of Language Users, «Handbook of Mathematical Psychology», v. 2, N. Y., 1963.
341. G. A. Miller and S. Isard, Free Recall of Self-embedded English Sentences, «Information and Control», No. 7, 1964.
342. E. A. Nida, The Identification of Morphemes, «Language», v. 24, No. 4, 1948.
343. E. Nida, Morphology. The Descriptive Analysis of Words, Ann Arbor, 1957.
344. A. G. Oettinger, Automatic Syntactic Analysis and the Pushdown Store, «Structure of Language and its Mathematical Aspects», N. Y., 1961.
345. Ch. Osgood, T. Sebeok, Psycholinguistics. A Survey of Theory and Research Problems, Baltimore, 1954.
346. H. Paul, Deutsche Grammatik, b. III, Halle, 1956.
347. K. L. Pike, Grammatical Prerequisites to Phonemic Analysis, «Word», v. 3, No. 2, 1947.
348. K. L. Pike, More on Grammatical Prerequisites, «Word», v. 8, No. 2, 1952.
349. K. L. Pike, Language in Relation to a Unified Theory of the Structure of Human Behaviour, Part I, Glendale, 1954.
350. H. Putnam, Some Issues in the Theory of Grammar, «Structure of Language and its Mathematical Aspects», N. Y., 1961.
351. R. S. Pittman, Nuclear Structures in Linguistics, «Readings in Linguistics» edited by M. Joos, Washington, 1957.
352. S. Saporita, Morph, Morpheme, Archimorpheme, «Word», v. 12, No. 1, 1956.
353. F. de Saussure, Mémoire sur le système primitif des voyelles dans les langues indo-européennes, Leipzig, 1879.
354. P. Schachter, Kernel and Non-kernel Sentences in Trans-

formational Grammar, «Preprints of Papers for the Ninth International Congress of Linguists», Cambridge, Mass., 1962.

355. T. A. Sebeok, The Informational Model of Language: Analog and Digital Coding in Animal and Human Communication, «Natural Language and the Computer», N. Y., 1963.

356. B. A. Sierstema, A Study of Glossematics; Critical Survey of the Fundamental Concepts, The Hague, 1955.

357. H. Spang-Hanssen, Typological and Statistical Aspects of Distribution as a Criterion in Linguistic Analysis, «Report for the Eighth International Congress of Linguists», v. I, Oslo, 1957.

358. H. Spang-Hanssen, Probability and Structural Classification in Language Description, Copenhagen, 1959.

359. R. P. Stockwell, The Transformational Model of Generative or Predictive Grammar, «Natural Language and the Computer», N. Y., 1963.

360. «Structure of Language and Its Mathematical Aspects», N. Y. 1961.

361. «Studia Grammatica», b. I, Berlin, 1962.

362. «Studia Grammatica», b. II, Berlin, 1963.

363. M. Swadesh, The Phonemic Principle, «Language», 1934, v. 10, No. 2.

364. H. Sweet, A New English Grammar, Logical and Historical, 1931.

365. L. Tesnière, *Éléments de syntaxe structurale*, Paris, 1959.

366. «The Prague School Reader in Linguistics», Prague, 1964.

367. K. Togeby, *Structure immanente de la langue française*, Copenhagen, 1951.

368. G. L. Trager, The Phonemes of Russian, «Language», v. 10, No. 4, 1934.

369. G. L. Trager, French Morphology: Verb Inflection, «Language», v. 31, No. 4, 1955.

370. G. L. Trager and H. L. Smith, *An Outline of English Structure*, Norman, 1951.

371. «Travaux linguistiques de Prague». *L'école de Prague d'aujourd'hui*, № 1, Prague, 1964.

372. W. F. Twaddell, On Defining the Phoneme, «Readings in Linguistics» edited by M. Joos, Washington, 1957.

373. «Universals of Language», ed. by J. H. Greenberg, Cambridge, Mass., 1963.

373a. D. Varga, Yngve's Hypothesis and Some Problems of the Mechanical Analysis, «Computational Linguistics», No. 3, Budapest, 1964.

374. J. T. Waterman, Perspectives in Linguistics, An Account of the Background of Modern Linguistics, Chicago and London, 1963.

375. R. Wells, Immediate Constituents, «Readings in Linguistics» edited by M. Joos, Washington, 1957.

376. R. Wells, De Saussure's System of Linguistics, «Readings in Linguistics» edited by M. Joos, Washington, 1957.

377. R. Wells, Meaning and Use, «Words», v. 10, No. 2—3, 1954.

378. R. W e l l s, Some Neglected Opportunities in Descriptive Linguistics, «Anthropological Linguistics», v. 5, No. 1, 1963.

379. D. S. W o r t h, The Role of Transformations in the Definitions of Syntagmas in Russian and Other Slavic Languages, «American Contributions to the V International Congress of Slavists», The Hague, 1963.

380. W. F. W y a t t J r., Structural Linguistics and the Laryngeal Theory, «Language», v. 40, No. 2, 1964.

381. V. H. Y n g v e, A Model and a Hypothesis for Language Structure, Technical Report 369, Cambridge, Mass., 1960.

382. V. H. Y n g v e, Random Generation of English Sentences, National Physical Laboratory, Teddington, Middlesex, 1961, paper 6.

383. V. H. Y n g v e, The Depth Hypothesis, «Structure of Language and its Mathematical Aspects», N. Y., 1961.

384. Y u e n-R e n-C h a o, The Non-uniqueness of Phonemic Solutions of Phonetic systems, «Readings in Linguistics» edited by M. Joos, Washington, 1957.

ОГЛАВЛЕНИЕ

Предисловие	3
Часть I. Из истории структурной лингвистики . .	7
Глава 1. Почему возникла структурная лингвистика	—
Глава 2. Непосредственные предшественники структурной лингвистики .	24
Глава 3. Классические школы структурной лингвистики	36
Часть II. Лингвистические модели	78
Глава 1. Понятие лингвистической модели	—
Глава 2. Типы лингвистических моделей .	99
Глава 3. Элементарные математические понятия	113
Часть III. Модели исследования	119
Глава 1. Модели дешифровки	120
Глава 2. Экспериментальные модели . . .	149
Часть IV. Модели речевой деятельности	197
Глава 1. Порождающие модели	198
Глава 2. Синтаксические модели анализа .	232
Глава 3. Семантическая модель анализа .	252
Часть V. Понятие метатеории	265
Заключение	280
Цитированная литература	283

Юрий Дереникович Апресян

Редактор *Г. В. Карпюк*

Художественный редактор

Н. А. Володина

Технический редактор

Н. Ф. Макарова

Корректоры *В. А. Глебова*

и *Н. И. Котельникова*

Сдано в набор 2/II 1966 г. Подписано к печати 11/XI 1966 г. 84×108^{1/2}₃₂. Печ. л. 9,5 (15,96). Уч.-изд. л. 16,42. Тираж 35 тыс. экз. (Тем. пл. 1966 г. № 89). А19002. Заказ № 571.

Издательство «Просвещение» Комитета по печати при Совете Министров РСФСР. Москва, 3-й проезд Марьиной рощи, 41.

Первая Образцовая типография имени А. А. Жданова Главполиграфпрома Комитета по печати при Совете Министров СССР. Москва, Ж-54, Валовая, 28.

Цена без переплета 66 к., переплет 10 к.

ЗАМЕЧЕННЫЕ ОПЕЧАТКИ

Страница	Строка	Напечатано	Следует читать
3	2 св.	XII века	VII века
50	3 сл.	etre	être
76	10 св.	или варианта	, или вариата
295	21 св.	Chacago	Chicago
295	4 сл.	fur	für
299	9 св.	Martiné	Martinet
299	11 св.	Martiné	Martinet

в кн. «Идеи и методы современной структурной лингвистики (краткий очерк)».