

И. И. РЕВЗИН

МОДЕЛИ ЯЗЫКА

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР

А К А Д Е М И Я Н А У К С С С Р

ИНСТИТУТ СЛАВЯНОВЕДЕНИЯ

И. И. РЕВЗИН

МОДЕЛИ ЯЗЫКА

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР

Москва 1962

ОТВЕТСТВЕННЫЙ РЕДАКТОР

В. Н. ТОПОРОВ

Памяти отца—геолога и страстного филолога-классика, погибшего в октябре 1941 года на фронте Отечественной войны, посвящаю эту книгу

От автора

Работы по машинному переводу, начавшиеся в Советском Союзе с 1955 г., явились у нас толчком для возникновения так называемой математической лингвистики, или, иначе говоря, применения математических идей и методов в языковедении.

Автор был привлечен к обсуждению лингвистических проблем машинного перевода А. А. Ляпуновым, беседы с которым оказали сильное влияние на всю последующую его деятельность. Автору посчастливилось одним из первых ознакомиться с теоретико-множественной системой грамматических понятий, предложенной ученицей А. А. Ляпунова О. С. Кулагиной, которая и в дальнейшем помогала автору в его работе над развитием и лингвистической интерпретацией этой теории.

Свои первые опыты в этом направлении автор изложил в мае 1957 г. на семинаре по математической лингвистике в МГУ. Участие в этом семинаре, работавшем с сентября 1956 г. по май 1958 г. под руководством Вяч. Вс. Иванова, П. С. Кузнецова и В. А. Успенского, помогло автору постепенно преодолеть первоначальное увлечение чисто внешней стороной математической символики и понять, что математические идеи в лингвистике плодотворны лишь там, где они связаны с ясным представлением о чисто лингвистической стороне тех или иных явлений. Участники семинара, несомненно, узнают в ряде параграфов идеи, обсуждавшиеся на семинаре.

В 1957/58 учебном году автор прочел на переводческом факультете 1-го МГПИИЯ спецкурс «Введение в математическую

лингвистику и машинный перевод». Вяч. Вс. Иванов и И. М. Яглом, ознакомившиеся со стенограммой курса, сделали ряд важных замечаний, использованных при написании настоящей книги.

Решающим обстоятельством, определившим структуру книги и ее теперешнюю направленность, явилось участие автора в работе сектора структурной типологии славянских языков Института славяноведения АН СССР. Товарищи автора по сектору М. И. Бурлакова, З. М. Волоцкая, Т. Н. Молошная, Т. М. Николаева, Д. М. Сегал, В. Н. Топоров и Т. В. Цивьян приняли живейшее участие в обсуждении книги и способствовали устранению некоторых из имевшихся в ней недостатков. Еще важнее было постоянное общение с сотрудниками сектора, уточнившее точку зрения автора на ряд типологических проблем.

С окончательным вариантом рукописи ознакомились также Вяч. Вс. Иванов, О. С. Кулагина, В. Ю. Розенцвейг, Г. А. Шестопал и И. Ю. Шехтер, сделавшие ряд чрезвычайно ценных замечаний.

Автор сожалеет, что не может упомянуть здесь всех тех, кто помог ему своими выступлениями в семинаре по математической лингвистике, в Комитете по прикладной лингвистике Секции речи Комиссии по акустике АН СССР, в объединении по машинному переводу, на 1-й Всесоюзной конференции по машинному переводу и других совещаниях и конференциях. В этих обсуждениях и частных беседах формировалась общая концепция теории моделей языка. Этим, разумеется, автор не снимает с себя ответственность за все недостатки этой общей концепции и все отдельные решения, предложенные им.

Рассматривая теорию моделей языка как лингвистическую теорию, автор основное место уделил не конструированию моделей, а выяснению связей между понятиями модели и лингвистическими фактами. Поэтому, между прочим, доказательство всех теорем вынесено в приложение. Поэтому же в основной части автор очень скупо пользуется математической символикой. Исключение сделано для некоторых теоретико-множественных символов, которые уже достаточно широко используются в языковедческих работах.

Тем, что в книге делается попытка построить именно лингвистическую, а не математическую теорию моделей, объясняется и ее структура. Так, отдельно в главе II разбираются фонологические модели, которые по своей структуре мало отличаются от некоторых грамматических моделей. Глава III специально посвящена обсуждению исходных понятий модели и их лингвистической интерпретации. Глава IV в основном посвящена обсуждению лингвистических приложений парадигматических моделей. Относительно синтагматических моделей (глава V) автору не удалось сделать это столь же полно, но и здесь по возможности указываются связи между рассматриваемыми моделями и соответствующими лингвистическими теориями.

Глава первая

ТИПЫ МОДЕЛЕЙ ЯЗЫКА

§ 1. Дедуктивные методы в лингвистике

В последнее время одним из основных вопросов, волнующих языковедов, стал вопрос о методах лингвистического исследования. Постановка этого вопроса вызвана не только возникновением ряда новых неожиданных приложений науки о языке (таких, как машинный перевод, проблема передачи и хранения информации и т. п.), но и внутренним развитием самого языковедения и свидетельствует, таким образом, об известной зрелости, достигнутой этой наукой.

В каждой науке наступает определенная эпоха, когда за периодом бурного развития, освоения все нового и нового материала, за рядом великолепных открытий возникает необходимость как бы вернуться назад и подвергнуть анализу самые основы этой науки, т. е. те фундаментальные понятия, которыми она оперирует. Наиболее характерно в этом отношении развитие математики — науки, с которой лингвистика в последнее время сближается теоретически и практически. Вот как описывает положение в математике к началу XIX в. проф. А. Я. Хинчин:

«Это была своеобразная картина: ни одно из самых основных понятий анализа не было определено сколько-нибудь точно, вопрос о том, что такое бесконечно малая величина, подвергался бесчисленным дискуссиям, с точки зрения логического обоснования совершенно бесплодным, так как в большинстве случаев ни одна из спорящих сторон не могла предложить ничего, кроме смутных, ни к чему не обязывающих образов»¹.

¹ А. Я. Хинчин. Краткий курс математического анализа. 2-е изд. М., 1955, стр. 613.

В дальнейшем ученые осознали, что без прочного логического фундамента математика как наука успешно развиваться не может. Огромные достижения современной математики, вплоть до создания электронных вычислительных машин, были бы немислимы без той ревизии логических основ науки, которая была проведена в XIX в.

Аналогичное состояние переживает и лингвистика. Увлеченные огромными достижениями исторического языкознания, связанными с применением сравнительно-исторического метода, ученые XIX в. почти не задумывались над точным значением терминов, таких, например, как «звук речи», «морфема», «слово», «предложение». Ф. де Соссюр был совершенно прав, говоря, что «до сих пор в области языка всегда довольствовались операциями над единицами, как следует не определенными»². Только глубокий переворот в языкознании, произведенный в первую очередь именно работами Ф. де Соссюра, заставил лингвистов по-настоящему задуматься над основными понятиями, которыми они оперируют.

Известно, что каждая наука пользуется абстракциями, однако пути, которые ведут к образованию этих абстракций, могут быть различными. Одни науки, например геология, ботаника, химия, физика, исходят из конкретных явлений, наблюдаемых в природе, и, обобщая наблюдаемые факты, идут от частного к общему; другие науки, такие как математика и логика, идут от общих абстрактных истин, взятых априорно, к установлению частных фактов. Конечно, не бывает чисто индуктивных или чисто дедуктивных наук. Так, химия и физика в их современном виде содержат весьма существенную дедуктивную часть, с другой стороны, даже в математике элемент индукции играет существенную роль³. Все же целесообразно выделять науки дедуктивные и науки индуктивные с точки зрения того, какие методы преобладают в данной науке. Куда же следует отнести языковедение с этой точки зрения? По самой своей природе языковедение должно прежде всего пользоваться индуктивными методами, оно описывает конкретные речевые факты конкретных языков, именно такая «полевая работа» лингвиста и представляет наибольшую ценность.

С другой стороны, наличие бесконечного множества речевых актов, изучаемых лингвистом, едва ли дает возможность сформулировать основные понятия науки о языке обобщением по индукции.

Отсюда следует, что лингвисты нуждаются не только в индуктивных, но и в дедуктивных методах исследования, чтобы

² Ф. де Соссюр. Курс общей лингвистики. Перевод с франц. М., 1933, стр. 111.

³ См., например: Д. ж. Пойя. Математика и правдоподобные рассуждения. Перевод с англ. М., 1958.

получить систему общих понятий, помогающих осмыслить те данные, которые добываются при анализе конкретных языков. При этом новые технические приложения языка требуют построения очень строгих дедуктивных систем, удовлетворяющих современным научным требованиям.

В своей дедуктивной части языковедение, по-видимому, может быть построено так, как строится логика или математика, а именно: выделяется некоторое минимальное количество первичных, неопределяемых терминов, а все остальные термины определяются через первичные. При этом должны быть четко сформулированы некоторые первичные утверждения о связи этих терминов между собой (аксиомы), и все остальные утверждения должны доказываться, т. е. сводиться к некоторым другим утверждениям.

Многочисленные споры вокруг математической лингвистики, имевшие место в последние годы, сводились в сущности к вопросу о том, может ли наука о языке быть построена таким способом. Не предвещая окончательного решения этого спора, которое принадлежит будущему, можно утверждать, что в настоящее время не ясно, каким образом уложить в рамки строгой дедуктивной теории все многообразие довольно противоречивых фактов речевой действительности, как они описаны в многочисленных грамматиках конкретных языков.

Такой вывод не дает, однако, повода для пессимизма. Современная наука обладает мощным орудием познания сложных явлений реальной действительности — методом моделирования, сущность которого заключается в том, что строится некоторая последовательность абстрактных схем, которые должны явиться более или менее близкой аппроксимацией данных конкретной действительности⁴.

Метод моделирования целесообразно применять в первую очередь к той части лингвистики, которую со времен Соссюра принято называть *н у т р е н и е й*. При этом необходимо сначала рассмотреть общие модели, не связанные с особенностями конкретных языков, что поможет в дальнейшем моделировать отдельные группы языков, а может быть, и конкретные языки. Итак, моделирование языка, как оно понимается в этой книге, есть метод, при котором исследователь исходит из некоторых наиболее общих черт конкретных языков, формулирует некоторые гипотезы о строении языка как абстрактной семиотической системы, а затем устанавливает, в каком отношении находятся следствия из этих гипотез и факты реальных языков, описываемые конкретными лингвистическими дисциплинами.

⁴ «Математическое представление (модель) языков отнюдь не тождественно самому языку» (см.: Н. Д. Андреев, Л. Р. Зиндер. Основные проблемы прикладной лингвистики. — ВЯ, 1959, № 4, стр. 18).

§ 2. Модель и интерпретация. Общий вид модели языка

Модель строится следующим образом. Из всего многообразия понятий, накопленных данной наукой, отбираются некоторые, которые удобно считать первичными. Фиксируются некоторые отношения между этими первичными понятиями, которые принимаются в качестве постулатов. Все остальные утверждения выводятся строго дедуктивно в терминах, которые определяют, в конечном счете, через первичные понятия.

Ясно, что термин «модель» в этом смысле абсолютно не соответствует эспировскому pattern. Модель в этом смысле не есть часть языка как системы, а представляет собой некоторое гипотетическое научное построение, некоторый конструкт⁵.

Необходимо строго различать модель, т. е. некоторый абстрактный объект, не зависящий от природы элементов, и ее лингвистическую интерпретацию (для разграничения понятий, относящихся к модели и ее лингвистической интерпретации, мы будем в дальнейшем употреблять два ряда разных терминов: для понятий модели — термины, принятые в математике, а при интерпретации — по возможности термины, принятые в лингвистике). Возможно несколько разных лингвистических интерпретаций одной и той же модели (например, фонологическая и грамматическая).

Взаимоотношения между моделью и ее интерпретациями сложны и многообразны. Модель, как уже говорилось, есть абстрактная конструкция, а «проблема интерпретации есть проблема подстановки чего-либо определенного»⁶. Для лингвистики это означает подстановку реальных лингвистических объектов вместо терминов модели. Заметим, что для теории лингвистических моделей особенно верно следующее замечание Б. Рассела:

«К вопросу об интерпретации незаслуженно относились с пренебрежением. Пока мы остаемся в области математических формул, все кажется определенным, но когда мы стараемся интерпретировать их, оказывается, что эта определенность в какой-то степени иллюзорна. Пока этот вопрос не выяснен, мы не можем сказать с какой-либо точностью, что, собственно, утверждает та или иная конкретная наука»⁷.

До тех пор, пока интерпретация той или иной лингвистической модели не исследована, сама модель остается чистой

⁵ О понятии «конструкт» см.: С. К. Шаумян. Операционные определения и их применение в фонологии. «Применение логики в науке и технике». М., 1960, стр. 150 и сл.

⁶ Б. Рассел. Человеческое познание, его сфера и границы. Перевод с англ. М., 1957, стр. 276.

⁷ Там же.

фикцией. В самом деле, доказательства, полученные дедуктивным путем, сколь бы безукоризненными они ни были с логической точки зрения, еще ничего не говорят о свойствах реального языка, описываемого моделью. Только в том случае, если выполняются определенные соответствия между оригиналом и моделью⁸, мы можем говорить о доказательности модели. Критерием этого соответствия может быть только практика, поэтому для теории моделей такую важность приобретает машинный перевод и другие практические приложения языкознания.

С другой стороны, исследование лингвистических интерпретаций модели чрезвычайно важно для выяснения условий более точного и полного совпадения модели с оригиналом, т. е. повышения познавательной ценности моделей.

Ниже мы увидим, какую роль в развитии теории моделей играет констатация неадекватности той или иной модели фактам конкретных языков.

Перейдем теперь к характеристике круга исходных первичных понятий, из которых строится большинство лингвистических моделей.

Прежде всего отметим, что в отличие, например, от акустической фонетики, которая имеет дело с определенным континуумом⁹, теория моделей языка всегда имеет дело с некоторыми дискретными единицами. На этом основании Джуз вообще считает лингвистику отраслью «дискретной математики»¹⁰. Как мы увидим ниже, теория моделей может рассматриваться как чисто лингвистическая теория, тем не менее осознание дискретного характера любой лингвистической модели чрезвычайно важно¹¹.

Во всех моделях языка в качестве исходного рассматривается понятие некоторого элемента. В зависимости от уровня языка, подлежащего моделированию, этот исходный элемент будет интерпретироваться как отдельный звук (на фонетическом уровне) или как словоформа (на уровне синтаксическом).

Вторым важным понятием является понятие кортежа. Мы называем кортежем любую упорядоченную последовательность элементов $A = x_1x_2...x_1...x_n$. Все кортежи мы разделяем на

⁸ См.: А. А. Зиновьев, И. И. Ревзин. Логическая модель как средство научного познания. «Вопросы философии», 1960, № 1, стр. 83—85.

⁹ См.: Н. С. Трубецкой. Основы фонологии. Перевод с нем. М., 1960, стр. 8—9.

¹⁰ М. J о о s. Description of language design. «Journal of the Acoustical Society of America», 1950, № 22, стр. 701—708.

¹¹ Ср. замечание Аристотеля о дискретном характере «речи»: А р и с т о т е л ь. Категории. Перевод А. В. Кубицкого. М., 1939, стр. 14.

два класса: кортежи первого класса мы будем называть отмеченными (в интерпретации это кортежи, принадлежащие данному языку), а кортежи второго класса — неотмеченными (это кортежи, не принадлежащие данному языку). Как и понятие «элемент», понятие «кортеж» получает разную интерпретацию в зависимости от уровня изучения языка. На фонологическом уровне под отмеченным кортежем мы понимаем «фонетическое слово». На синтаксическом уровне мы интерпретируем кортеж как фразу. Поскольку, однако, понятие отмеченной фразы допускает различную интерпретацию, мы просто считаем, что рассматриваемые в некоторой модели фразы нам уже заданы извне.

Мы подчеркиваем это обстоятельство, употребляя термин «отмеченные фразы».

Естественно разбить все модели языка на четыре основных вида в зависимости от того, является ли число элементов конечным, а длина отмеченного кортежа ограниченной¹²:

- 1) число элементов конечно, и длина отмеченного кортежа ограничена;
- 2) число элементов конечно, и длина отмеченного кортежа не ограничена;
- 3) число элементов бесконечно, и длина отмеченного кортежа ограничена;
- 4) число элементов бесконечно, и длина отмеченного кортежа не ограничена.

Каждый из перечисленных видов моделей ставит перед исследователем свои логические вопросы. Мы будем рассматривать лишь модели, где множество элементов конечно. Что же касается длины кортежа, то сначала мы не будем накладывать на нее ограничений, а затем посмотрим, какова разница между моделью первого класса и моделью второго класса.

Существенным исходным (иногда производным) понятием модели языка является понятие разбиения множества элементов на подмножества. Иначе говоря, обычно считается заданной некоторая система подмножеств исходного множества, и для каждого элемента указано, к каким множествам он принадлежит. Эти подмножества могут, вообще говоря, пересекаться, таковы признаки в фонологии или категории в грамматике.

Большинство моделей строится так, чтобы, основываясь на этом исходном разбиении, получить некоторое разбиение на непересекающиеся классы. Иногда некоторое разбиение на непересекающиеся классы задается дополнительно. Так или иначе все модели, рассматриваемые нами, сводятся к намеченному здесь кругу понятий элемента, кортежа и некоторых разбиений множества исходных элементов.

¹² См.: J. Greenberg. Essays in linguistics. Chicago, 1958.

§ 3. Аналитические и синтетические модели

Логическая структура каждой модели такова, что мы в ней имеем дело с некоторыми формальными операциями над некоторым множеством объектов. По отношению к операциям, допустимым в данной модели, все объекты делятся на три вида: а) исходные объекты, к которым могут применяться некоторые операции, но которые сами не возникают в результате каких-то операций (а заданы извне); б) промежуточные объекты, к которым применяются операции и которые сами возникают в результате применения операций; в) заключительные объекты, которые возникают в результате операций, но к которым неприменимы операции.

Мы будем говорить, что модель А есть обратная модель по отношению к модели В, если объекты, являющиеся исходными в А, суть заключительные объекты в В и обратно, исходные объекты В суть заключительные объекты А. Построение модели, обратной по отношению к А, мы условно назовем «обращением» модели А.

Мы разобьем модели языка на аналитические (или «анализирующие»¹³) и, обратные по отношению к ним, синтетические (или «синтезирующие») в зависимости от того, исходим ли мы из множества отмеченных кортежей (аналитическая модель) или получаем отмеченные кортежи в результате некоторых операций (синтетическая модель, или, как иногда говорят, модель порождения). Мы будем каждый раз начинать с рассмотрения некоторых аналитических моделей и тут же указывать на соответствующую синтетическую. Аналитические модели в известной степени формализуют ситуацию, рассматриваемую в дескриптивной лингвистике (например, понятию множества отмеченных кортежей соответствует «universe of discourse» у Харриса¹⁴). Грамматики порождения¹⁵ стали строиться в последнее время, и на базе понятий дескриптивной лингвистики.

Эти два вида моделей, аналитические и синтетические, соответствуют двум возможным видам лингвистического описания, а именно: один путь — от речевых фактов к системе языка, а второй — от системы языка к речевым фактам. В какой-то мере они соответствуют и двум аспектам акта коммуникации: слушанию («анализ») и говорению («синтез»), причем неоднократно отмечалось, что этим аспектам должны соответствовать

¹³ См.: Г. С. Це й т и н. К вопросу о построении математических моделей языка. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста», вып. 3. М., 1961.

¹⁴ Z. H a r r i s. *Methods in structural linguistics*. Chicago, 1951, стр. 7—9.

¹⁵ См.: Н. Х о м с к и й. Синтаксические структуры. «Новое в лингвистике», вып. II, М., 1962, стр. 412 — 527.

разные типы описания (ср. различие между «морфологией» и «синтаксисом» у Есперсена¹⁶, между «пассивной» и «активной» грамматикой у Щербы¹⁷, между «языковой дешифрацией» и «языковой стилизацией» у Матезиуса¹⁸, указание на разный характер фонологических единиц, соответствующих обоим аспектам, у Якобсона¹⁹). Близко делению на аналитические и синтетические модели и различие между двумя подходами к лингвистическому анализу у Хоккета²⁰: модели, в которых мы исходим из некоторых единиц и их взаимоотношений в тексте (*item and arrangement*) и модели, в которых мы исходим из некоторых единиц и некоторого процесса, позволяющего получать новые единицы (*item and process*).

Один из кардинальных фактов теории лингвистических моделей — следующий: одно и то же множество отмеченных кортежей может быть порождено двумя различными синтетическими моделями (простейший пример такой неоднозначности, приведенной Гринбергом²¹, дан в приложении). Этот факт говорит о том, что один и тот же язык (как некоторое множество отмеченных кортежей) может быть описан несколькими разными грамматиками. Совокупность отмеченных кортежей, которую можно описать с помощью различных правил порождения, Гринберг предлагает называть «гетеронимической знаковой системой». «Гетеронимия» в этом смысле и есть обобщение явлений омонимии на случай, когда в качестве единицы выступает целая знаковая система (пример гетеронимии дан в приложении).

В примере Гринберга гетеронимия возникает из-за неопределенности исходных единиц и допустимых операций. Однако логически гетеронимия не исключена и при более определенном задании модели. Поэтому целесообразно для всякой аналитической модели указывать соответствующие ей синтетические. Вообще говоря, явления гетеронимии (так же как и некоторые другие факты, с которыми мы познакомимся позднее) показывают, что синтетические модели более удобны для описания языка, чем модели аналитические. С другой стороны, мы уви-

¹⁶ О. Есперсен. *Философия грамматики*. Перевод с англ. М., 1958, стр. 39—46.

¹⁷ Л. В. Щерба. *Преподавание иностранных языков в средней школе*. М.—Л., 1947, стр. 84 и сл.

¹⁸ V. M a t h e s i u s. *Obsahový rozbor současné angličtiny na základě obecně lingvistického*. Praha, 1961, стр. 11—13.

¹⁹ R. J a k o b s o n and M. H a l l e. *Fundamentals of language*. 's-Gravenhague, 1956.

²⁰ C h. H o c k e t t. *Two models of grammatical description*. «Word», 1954, № 10.

²¹ J. H. G r e e n b e r g. Указ. соч., стр. 7. Пример Гринберга тривиален, хотя и нагляден. В лингвистике на возможность двойного описания одного и того же явления уже обращали внимание, ср.: J u e n - R e n C h a o. *The non-uniqueness of phonemic solutions of phonetic systems*. «Readings in linguistics». Ed. by M. Joos. New York, 1958, стр. 38—54.

дим, что построение разумных синтетических моделей всегда предполагает наличие хорошо разработанных аналитических моделей. Совместное рассмотрение аналитических и синтетических моделей важно и с точки зрения интерпретации модели и оценки степени ее адекватности.

Приемлемость аналитической модели может оцениваться с той точки зрения, насколько сформулированные в ней понятия отражают системные отношения в языке. Однако известно, что именно описание системных отношений в языке связано со значительными трудностями и далеко не однозначно.

Приемлемость синтетической модели оценить легче: мы рассматриваем множество кортежей, порожденной данной моделью, и сопоставляем его с множеством фраз, фактически существующих в данном языке. Если одна модель производит больше отмеченных фраз и меньше неотмеченных фраз²², чем другая, то она, разумеется, лучше. Если две модели порождают одинаковое множество фраз, то их можно было бы считать эквивалентными и т. п. Исходя из этих соображений, Н. Хомский в указанной работе построил ряд моделей, таких, что каждая последующая модель производит все фразы предыдущей, но обратное не имеет места.

В общем виде этот вопрос был поставлен в упомянутой работе Г. С. Цейтина, который пишет: «Можно уточнить задачу аппроксимации языка моделями, вводя понятие сходимости последовательности моделей к языку, которое, скажем, при моделировании множества грамматически правильных фраз могло бы определяться следующим образом: говорим, что последовательность моделей сходится к языку, если каждая грамматически правильная фраза будет грамматически правильной и во всех моделях последовательности, начиная с некоторого места, и каждая последовательность знаков, не являющаяся грамматически правильной фразой, не будет такой и во всех моделях последовательности, начиная с некоторого места»²³.

²² В работе: F. W. H a r w o o d. Axiomatic syntax («Language», vol. 31, 1955, № 3) предложены (для первого вида из § 2) следующие оценки эффективности модели. Пусть L — множество отмеченных кортежей, а M — множество неотмеченных кортежей. Пусть C — множество порождаемых моделью кортежей, а D — дополнение к нему. Пусть LC есть пересечение L ∩ C, а MD пересечение M ∩ D. Обозначим через n(M), n(L), n(LC) и n(MD) — число кортежей в соответствующем множестве. Тогда можно ввести следующие оценки

$$F = \frac{n(LC)}{n(L)} \quad \text{и} \quad f = \frac{n(MD)}{n(M)}.$$

Идеальным является случай, когда $F = f = 1$, т. е. когда C совпадает с L. Отрицательная оценка не имеет смысла, когда $n(M) = 0$, т. е. любые кортежи отмечены. Крайним является также случай $F = 1$, $f = 0$, т. е. когда не все кортежи отмечены, но все они порождаются данной моделью.

²³ Г. С. Цейт и н. Указ. соч., стр. 3.

Заметим в заключение, что выбор модели зависит и от характера исследуемого материала. Если множество отмеченных фраз бесконечно, то аналитическая модель играет лишь вспомогательную теоретическую роль, помогая выяснению понятий, но не может быть непосредственно применена. Если же множество отмеченных фраз конечно, то аналитическая модель, по-видимому, удобней для практического применения, чем синтетическая. Мы вернемся к этому вопросу в § 45, здесь же отметим, что разница в лингвистическом подходе к конечному или бесконечному множеству фраз именно в связи с анализом или синтезом была отмечена еще Л. В. Щербой, который писал: «В мертвых языках... в опыте не дано никаких процессов говорения и понимания, зато имеется материально-объективированный языковой материал в виде текстов... Указанным здесь различием между мертвыми и живыми языками объясняется, между прочим, относительная легкость исследования первых, где имеется конечное количество объективированного языкового материала, и большая трудность исследования вторых, где материал бесконечен и трудно объективируется»²⁴.

§ 4. Распознающие модели

Существует особый тип моделей, в котором считаются заданными и множество отмеченных кортежей и система порождения и рассматривается процесс перехода от кортежей к системе, а именно исследуются способы такого перехода в минимальное число шагов.

Наиболее подробно модель такого типа применительно к русскому языку описана Т. М. Николаевой²⁵.

Ясно, что модели этого типа имеют очень большое прикладное значение (машинный перевод, обучение иностранным языкам и т. п.).

Построение распознающих моделей имеет и глубокий теоретический смысл: ведь суть такой модели в том, что при анализе восстанавливаются единицы синтеза. Это соответствует новейшим кибернетическим представлениям о природе распознавания речи, развиваемым в последнее время Л. А. Чистович²⁶. Согласно этим представлениям при анализе речи необходимо восстановить те инструкции, которые привели к порождению данного речевого отрезка.

²⁴ Л. В. Щ е р б а. Указ. соч., стр. 74.

²⁵ Т. М. Н и к о л а е в а. Опыт построения алгоритмической морфологии русского языка. «Структурно-типологические исследования», вып. 1 (в печати).

²⁶ Л. А. Ч и с т о в и ч. Текущее распознавание речи человеком. «Машинный перевод и прикладная лингвистика», 1961, № 6.

§ 5. *Парадигматические и синтагматические модели*

Модели языка можно классифицировать и с другой точки зрения, а именно с точки зрения тех отношений в языке, которые отражаются в них. Условно одни из них можно назвать парадигматическими, а другие синтагматическими. К парадигматическим относятся те модели, в которых исследуются принципы объединения некоторых элементов в классы (отдельных звуков в фонемы, отдельных морф в морфемы, отдельных слов в категории и части речи и т. п.) и установления отношений в системе. К синтагматическим относятся модели, в которых исследуются отношения между элементами (фонемами, словами) в некотором фиксированном corteже, т. е. в речи. Как известно, оба аспекта (синтагматика и парадигматика) в языке теснейшим образом связаны между собой. Так, с одной стороны, делаются попытки вывести парадигматические отношения из синтагматических. С другой стороны, удобнее строить синтагматические модели после того, как элементы уже объединены в классы, ибо это упрощает описание.

Важность парадигматических моделей в том, что они позволяют моделировать системные отношения в языке. На почве парадигматических моделей возникают модели динамические, рассматривающие язык как систему, переходящую от одного состояния устойчивости к другому. Очевидно, что именно такие модели необходимы для внедрения кибернетических идей в сравнительно-историческое изучение языков.

§ 6. *Модели языка и статистика речи*

Метод моделирования как метод дедуктивного анализа языка может оказать помощь и проникновению в языковедение точных индуктивных методов, а именно методов статистических.

В свою очередь моделирование языка должно быть увязано с учетом его статистической структуры. Это вытекает из следующих соображений.

Прежде всего, модель имеет в виду некоторый идеальный объект, соответствие этого идеального объекта реально существующим соотношениям всегда более или менее относительно и степень этого соответствия должна выясняться статистически. Далее мы увидим, что всюду, где происходит переход от определенных абстрактной модели к применению этих определений для анализа реального языка, явно или неявно появляются статистические критерии. Да это и не может быть иначе. Те или иные категории в языке выделяются, в сущности, на основании статистических критериев. Если модель должна близко описывать реальную действительность языка, то выделяемые

в ней группировки должны соответствовать тем, которые в языке выделяются статистически.

Во-вторых, любая модель явно или неявно использует понятие исходного речевого материала (некоторого множества слов или фраз). Этот исходный речевой материал всегда составляет некоторую часть языка. Даже если модель имеет в виду бесконечное множество кортежей, то при переходе от модели к анализу конкретного материала мы снова сталкиваемся с некоторым конечным множеством, или, говоря статистически, с некоторой «выборкой». Вопрос о том, какой должна быть величина выборки — типично статистический вопрос.

В-третьих, в моделях фигурируют такие логические термины, как «все», например «все фразы вида...». Как только мы переходим к применению модели, то коррелятом этого термина становится статистический термин «достаточно большое число» (ср. § 45).

В-четвертых, когда мы имеем некоторый ряд моделей, описывающих один и тот же объект, то возникает вопрос об оценке их эффективности. Этот вопрос имеет кроме чисто логической стороны (ср. § 3) очень важный статистический аспект.

В-пятых, основное понятие модели, а именно «отмеченности фразы», должно также интерпретироваться на основании статистических соображений. В силу своеобразной «презумпции осмысленности» всегда имеется вероятность того, что фраза, которую мы считаем «неотмеченной», будет каким-то числом людей интерпретироваться как осмысленная и лишь после значительного числа экспериментов можно приписать каждой фразе вероятность быть осмысленной²⁷.

Правда, в данной работе вопросы лингвистической статистики специально разбираться не будут, но о существующей между моделированием и статистикой связи всегда следует помнить.

§ 7. Теория моделей и структурная типология языков

В работах, посвященных современному состоянию типологических исследований²⁸, подчеркивается важность новых точных методов для типологии языков. Особое место здесь, по-ви-

²⁷ См. по этому поводу справедливые возражения Хилла о критерии «грамматичности» Хомского: A. A. Hill. Grammaticality. «Word», vol. 17, 1961, № 1, стр. 1—10.

²⁸ См.: Вяч. Вс. Иванов. Типология и сравнительно-историческое языкознание.— ВЯ, 1958, № 5; V. Skalička. O současném stavu typologie.— SaS, 1958, № 3; е г о ж е. Z nové typologické literatury.— SaS, 1960, № 1.

димому, принадлежит методу моделирования²⁹. Мы уже говорили, что любая модель отличается от некоторого реального языка, что точно описать язык можно лишь некоторой последовательностью моделей, может, быть даже бесконечной. Чрезвычайно важно, однако, что разные языки по-разному отличаются от некоторой данной модели. Тем самым модель становится объективным эталоном для сравнения между собой отдельных языков.

Этот метод может оказаться, по-видимому, особенно полезным для изучения близкородственных языков, например славянских, поскольку здесь многие критерии, до сих пор рассматривавшиеся в типологии и связанные, например, со структурой слова, приводят к тривиальным результатам. В дальнейшем мы на ряде примеров убедимся в том, что моделирование помогает типологическому сравнению языков.

По-видимому, в дальнейшем теория моделей будет увязана с типологией языков и по другой линии, а именно с точки зрения построения общей типологии знаковых систем. Речь идет о замечательной идее, высказанной А. А. Зализняком, Вяч. Вс. Ивановым и В. Н. Топоровым и состоящей в том, что каждый язык можно рассматривать как моделирующую систему, в связи с чем может быть построена целая иерархия моделей, начиная от абстрактных, с сравнительно малой моделирующей способностью, но зато с более широким охватом материала (именно модели такого типа и рассматриваются в последующих главах), и кончая моделями с высокой моделирующей способностью, но с весьма ограниченной предметной областью.

При такой постановке проблемы, интересующей нас в этой работе, вопрос о соотношении между моделью и ее лингвистической интерпретацией есть вопрос о переходе с одного уровня моделирования на другой.

§ 8. Теория моделей как лингвистическая теория

Теория моделей языка, по-видимому, представляет чисто логический и, может быть, даже математический интерес. Так, относительно моделей, рассматриваемых в главе V данной книги, И. Бар-Хиллел утверждает, что «существуют крайне интересные связи между теорией лингвистических моделей и такими теориями, как теория автоматов, теория рекурсивных функций (возможно, особенно полезная в том виде, какой она приняла в теории алгоритмов) и теория канонических систем Поста.

²⁹ См.: М. И. Бурлакова, Т. М. Николаева, Д. М. Сегал, В. Н. Топоров. Структурная типология и славянское языковедение. «Структурно-типологические исследования», вып. 1.

Эти многочисленные связи показывают, что в теории языковых моделей мы имеем, по-видимому, интересное новое поле взаимного оплодотворения математической логики и структурной лингвистики, которое может привести к важным результатам»³⁰.

Однако в нашей работе теория лингвистических моделей рассматривается не как математическая или логическая, а как лингвистическая теория.

А именно теория моделей языка есть с точки зрения автора часть общего языкознания, которая использует построенные в тех или иных целях математические модели языка для:

а) выявления круга исходных понятий, которыми оперируют языковеды, изучая язык, и установления отношений между выделенными единицами (ср. круг проблем, связанных с построением так называемой аксиоматики языкознания);

б) выяснения приемов, параллельно применяемых под разными названиями на разных уровнях исследования языка (ср. круг проблем, связанных с понятием «изоморфизма» в лингвистике);

в) выяснения связи между отношениями в тексте и отношениями в системе (ср. круг проблем, поднимаемых в так называемой глоссематике).

С другой стороны, теория моделей языка должна помочь установить, какие из способов описания языка, предложенные лингвистами, могут быть формализованы уже сегодня, а какие требуют создания новых моделей, способных отразить более сложные построения, применяемые в лингвистике.

³⁰ Y. Bar-Hillel. Decision procedure in Natural Languages. «Logique et analyse», Nouvelle série, 2 an., Janvier 1959, № 5, стр. 29.

МЕТОДЫ МОДЕЛИРОВАНИЯ
В ФОНОЛОГИИ

**§ 9. Основные понятия при построении
фонологических моделей и их лингвистическая
интерпретация**

При исследовании фонетических явлений речи и их дальнейшей фонологической интерпретации мы имеем дело со следующими исходными объектами:

1) Нам заданы некоторые исходные элементы, а именно звуки речи¹.

2) Задана некоторая совокупность фонетических категорий, или признаков, а именно: звонкость, глухость, мягкость, твердость, ффрикативность, смычность, открытость, закрытость² и т. п.

Каждый звук речи поставлен в соответствие с некоторым подмножеством множества признаков. Используя более обычную терминологию, мы будем говорить, что каждый звук состоит из *n* признаков.

3) Заданы некоторые отмеченные кортежи звуков речи, которые интерпретируются как «фонетические слова».

Надо сразу же заметить, что подобная интерпретация связана со значительной трудностью. Дело в том, что возможны два принципиально разных понимания фонетического слова.

Интерпретация 1. Под фонетическим словом понимается

¹ Об этом понятии см.: П. С. Кузнецов. Об определении фонемы. «Бюллетень объединения по проблемам машинного перевода», № 5. М., 1957.

² Для простоты мы будем пользоваться привычными артикуляционно-перцептивными признаками, а не акустическими признаками, описанными в работе: Р. Якобсон, Г. М. Фант и М. Халле. Введение в анализ речи, «Новое в лингвистике», вып. II, М., 1962, стр. 173—230; в некоторых случаях мы будем указывать, какие изменения вносятся при обращении к этой системе признаков.

минимальный отрезок между двумя паузами в реально существующем речевом отрезке данного языка. Эта интерпретация обычно и имеется в виду в фонологии, где, например, основной критерий выделения единиц состоит в сравнении двух реально существующих фонетических слов (или даже лексических слов).

Интерпретация 2. Фонетическим словом называется минимальное сочетание звуков речи, допустимое в данном языке между двумя паузами. Понятие допустимости, разумеется, недостаточно формально. Неясно, в какой мере можно включить сюда случаи «заумного языка», например:

Дыр бул шыл
Убецур
(Крученых).

Вообще говоря, надо иметь в виду, что для построения стихотворения на заумном языке поэт берет не любые допустимые сочетания, а, как правило, те, которые избегаются в естественной речи³, ср.:

Бобзоби пельсь губы
Взоеми пельсь взоры
(Хлебников)

Мы будем называть сочетание звуков допустимым в данном языке, если каждый элемент этого сочетания встречается в некотором реально зафиксированном слове данного языка в таком же окружении, что и в данном сочетании. По этому определению отрывок из Крученых состоит из фонетических слов, а *взоеми* не является фонетическим словом русского языка.

Подобное определение позволит в дальнейшем легко переходить от интерпретации 1 к интерпретации 2 (слова «хотя бы одно фонетическое слово по интерпретации 1» соответствуют словам «любое фонетическое слово по интерпретации 2»).

Интерпретация 2 удобнее, чем первая, для моделирования по следующим соображениям:

1) понятие лексического и грамматического значения нецелесообразно вводить в рассмотрение, поскольку осложняется исходная ситуация, а основная разница между фонетическим словом в одной и другой модели в том, что в интерпретации 2 слова не обязательно имеют лексическое и грамматическое значение;

2) фонолог обязан иметь в виду не только реализованные в словах языка, но потенциальные сочетания звуков речи. Как известно, такие сочетания резко различаются от языка к языку

³ Ср. статьи Л. Якубинского: «О поэтическом глоссемосочетании» и «Скопление одинаковых плавных в практическом и поэтическом языках». — «Поэтика. Сборники по теории поэтического языка», т. 1. Пг., 1919.

и поэтому представляют самостоятельный интерес. Кроме того, обогащение словаря путем не книжного заимствования, как правило, происходит так, что подбирается потенциально возможное сочетание звуков речи одного языка, более или менее похожее на слово другого ⁴;

3) при построении порождающей модели, как мы увидим в § 16, порождение только фонетических слов в интерпретации 1 невозможно;

4) при моделировании полезно, чтобы явления фонологии и грамматики описывались более или менее параллельно. Между тем описание множества отмеченных кортежей в § 21 соответствует именно интерпретации 2.

Исходя из сказанного, мы будем иметь в виду интерпретацию 2 исходного материала. Если же для более легкого восприятия мы будем иногда пользоваться интерпретацией 1, то будут строиться лишь такие определения, для которых в силу сделанного выше замечания интерпретация 1 легко сводится к интерпретации 2.

Необходимо учитывать, что выбор интерпретации 2 в качестве основной ограничивает наши возможности. Мы уже не сможем в дальнейшем определить «фонему» как мельчайшую единицу, служащую для смыслоразличения. Можно показать, однако, что такое определение приводит к ненужным формальным затруднениям. С другой стороны, можно определить, исходя из явлений дистрибуции, на заданном нам исходном материале некоторую единицу, достаточно близкую к тому, что обычно понимают под фонемой. Эта весьма плодотворная — с точки зрения дальнейшей формализации — концепция фонемы была впервые предложена Джоунзом, который писал: «Если два звука языка могут встречаться в одной и той же позиции по отношению к окружающим звукам, то они по определению относятся к разным фонемам» ⁵.

Именно это понятие фонемы и ляжет в дальнейшем в основу построения фонологических моделей в данной работе. Поскольку фонология является наиболее развитой абстрактной областью изучения языка, то читатель не должен ожидать здесь чего-либо существенно нового по сравнению с трудами классиков фонологии. Данная глава должна подготовить, на более знакомом материале, к восприятию материала следующих глав. Тем не менее автор надеется, что ряд формулировок будет воспринят как полезное уточнение фонологических понятий.

⁴ Ср. описание этого процесса у Л. Толстого в «Войне и мире» (т. IV): «*Vive Henri Quatre! Vive ce roi vaillant!* — пропел Морель, подмигивая глазами. — *Ce diable à quatre.*... — *Виварика! Виф серуару! сидяблика.*... — повторил солдат, взмахнув рукой и действительно уловив напев». Интересно, что здесь именно русские фонетические слова с оглушением звонких в абсолютном исходе (*vif*) и с переходом *e* в *i* в предударной позиции.

⁵ D. J o n e s. On phonemes. — TCLP, Prague, 1931, № 4, стр. 77.

§ 10. Две классификации фонетических признаков. Фонема и архифонема

Фонетические признаки можно классифицировать по отношению к изолированным звукам и звукам в потоке речи. Обратимся прежде всего к отдельным звукам. Пусть анализируется некоторый язык, заданный тремя исходными понятиями из § 9.

Мы будем называть два признака совместимыми в данном языке, если существует хотя бы один звук речи, в котором они выступают вместе (иначе говоря, если оба входят в подмножество признаков, поставленное в соответствие с некоторым звуком речи; в дальнейшем читатель может сам в случае необходимости переводить все формулировки этого параграфа на язык соответствий).

В противном случае мы будем говорить о несовместимых признаках. Примеры совместимых признаков в русском языке: глухость и фрикативность, звонкость и мягкость и т. п. Примеры несовместимых признаков: глухость и звонкость, мягкость и твердость, твердость и открытость. Несовместимые признаки могут быть однородными и неоднородными.

Мы будем называть два несовместимых признака m_1 и m_2 однородными, если существует хотя бы один звук данного языка, такой, что замена m_1 на m_2 (или m_2 на m_1) дает также некоторый звук данного языка.

Примеры однородных признаков в русском языке: глухость и звонкость; верхний подъем, средний подъем и нижний подъем; мягкость и твердость. Неоднородными несовместимыми признаками являются, например, твердость и открытость.

Примечание. В упомянутой выше (см. сноску 2 на стр. 20) системе акустических признаков Р. Якобсона, по-видимому, всякие два несовместимые признака однородны.

Обратимся теперь к классификации признаков в потоке речи. Мы условились, что каждый звук может быть представлен в виде совокупности признаков. Таким образом, каждое фонетическое слово есть упорядоченная последовательность совокупностей признаков. Для общности будем считать, что пауза есть также звук речи, поставленный в соответствие с признаком «молчание» (или состоящий из одного признака «молчание»).

Рассмотрим некоторое допустимое в данном языке сочетание звуков (наиболее удобно было бы интерпретировать «допустимое сочетание» как «слог», но это понятие, к сожалению, трудно поддается моделированию ⁶).

⁶ Необходимо, однако, указать на интересные работы Л. Ельмслева в этом направлении: «La syllabation en slave» («Зборник у част Белића».

Если в данной допустимой паре звуков S_1S_2 один из признаков (пусть это будет признак m) звука S_1 (или S_2) нельзя заменить никаким другим так, чтобы вновь получилось допустимое сочетание звуков, то мы будем говорить, что данный признак m связан в данной паре. Возьмем теперь звук S_1 и рассмотрим множество всех пар, в которых этот звук стоит на данном месте (например, первом). Если во всех этих парах признак m связан, то мы будем называть этот признак *нерелевантным*. Остальные признаки мы будем называть *релевантными* (очевидно, что так определенный «релевантный признак» — понятие более широкое, чем «дифференциальный признак»).

В качестве примера рассмотрим звуки ε и e в русском языке. Они различаются, как известно, лишь одним признаком: открытостью или соответственно закрытостью. Возьмем теперь все пары с e на первом месте, например *et'* (в *эту*), *ed'* (в *веду*), *en'* (в *тень*) и т. п. Во всех этих парах признак закрытости связан (нельзя заменить признак «закрытость» признаком «открытость», не изменив и второго звука пары).

Если взять все пары, в которых встречается ε , то опять-таки замена его на e приводит к недопустимому сочетанию. Поэтому мы считаем оба признака нерелевантными.

Для того чтобы было видно, что наш принцип не отличается от принятого в фонологической литературе, мы ниже почти в тех же словах повторим рассуждение, проводимое Р. И. Аванесовым⁷. Возьмем слова *s'at'* и *s'n'at'*. Звук *s'* характеризуется признаками: зубной, щелевой, глухой, мягкий. Первые три признака не связаны в парах, *s'ā* и *s'n'* (ср. *p'at'*, *p'n'ā*, *gāt'*, *agn'ā*, *z'āt'*, *maz'n'ā*). Что касается признака «мягкость» в звуке *s'*, то в паре *s'ā* он не связан (ср. *pisāt'*), а в паре *s'n'* этот признак связан, так как пара *sn'* недопустима. Итак, признак мягкости у *s'* в одних парах не связан, а в других связан; поэтому он считается релевантным.

Если некоторый релевантный признак связан в данном сочетании, то мы будем называть данное сочетание *позитивным* и *нейтрализацией* противопоставления данного признака и всех однородных с ним.

Например, позиция перед паузой в русском языке есть позиция нейтрализации противопоставления признаков звонкости и глухости.

Наконец, мы будем называть *фонемой* любую совокупность релевантных неоднородных признаков, поставленных в соответствие с некоторым звуком речи. Два звука,

Београд, 1937) и «The syllable as a structural unit» (Proceedings of the Third International Congress of phonetic science).

⁷ Р. И. Аванесов. Кратчайшая звуковая единица в составе слова и морфемы. «Вопросы грамматического строя». М., 1955, стр. 119.

соответствующих одной фонеме, называются двумя а л л о ф о н а м и, или в а р и а н т а м и одной фонемы.

По-видимому, то, что понимается под фонемой в школе Л. В. Щербы, в особенности у его учеников⁸, практически совпадает с тем, что предусмотрено нашей моделью.

В дальнейшем нам понадобится еще одно понятие, а именно понятие «архифонемы». Под архифонемой обычно понимается совокупность релевантных признаков, общих для двух фонем⁹.

Пусть имеется две разных фонемы, каждая из которых содержит n признаков. Пусть у них $n-1$ признаков совпадают. Тогда мы будем называть эту совокупность $n-1$ признаков а р х и ф о н е м о й.

Обычно рассматриваются лишь такие архифонемы, для которых хотя бы один из признаков, различающих две фонемы, имеет позицию нейтрализации. Например, рассматривается архифонема «неносовой, дентальный, смычный, твердый» как совокупность признаков общих для пары фонем d и t в русском языке, поскольку однородные признаки звонкость и глухость имеют позиции нейтрализации.

В нашей модели мы будем пользоваться для рассмотренного случая термином «актуальная архифонема». Однако в общем случае мы будем просто говорить об архифонемах, не уточняя, являются ли они актуальными или нет. Если две фонемы имеют общую актуальную архифонему, т. е. различаются только тем, что в одной присутствует признак m_1 , а в другой признак m_2 , то мы будем говорить, что они входят в одну элементарную фонологическую категорию.

§ 11. Фонема в московской фонологической школе и ее формальный аналог

Существует еще одна важная единица фонологической системы, которая рассматривается московской фонологической школой и также называется в ней «фонемой» или, как более удачно выражается Р. И. Аванесов, «фонемным рядом»¹⁰. Этим последним термином мы и будем пользоваться в дальнейшем. Необходимо, однако, помнить, что, собственно говоря, мы хотим построить лишь формальный (дистрибутивный) аналог для «фонемного ряда».

Введем прежде всего одно вспомогательное понятие. Мы будем говорить, что фонема F_1 доминирует над фонемой F_2 , если существует хотя бы одно фонетическое слово, в котором F_1 можно заменить на F_2 , и хотя бы одно фонетическое слово, в котором F_2 нельзя заменить на F_1 .

Так в русском языке o можно заменить на a в слове *godu* (годы), полу-

⁸ См.: Л. Р. Зиндер. Специфические особенности восприятия звуков речи. «Восприятие звуковых сигналов в различных акустических условиях». М., 1956, в особенности стр. 70.

⁹ См.: Н. С. Трубецкой. Основы фонологии. М., 1960, стр. 87.

¹⁰ Р. И. Аванесов. Фонетика современного русского литературного языка, М., 1956, стр. 31 и сл.

чив *gady* (*гады*), но *a* нельзя заменить на *o* в слове *vada* (*вода*). Поэтому *o* доминирует над *a*. Другой пример: *g* можно заменить на *k* в слове *gara* (*гора*), получив *kara* (*кора*), но *k* нельзя заменить на *g* в слове *pirok* (*пирог*). Поэтому *g* доминирует над *k*. Очевидно, что *d* также доминирует над *k* (ср. *dom* и *kot*, но только *pirok*). В связи с этим введем новое определение.

Мы будем говорить, что фонема F_1 абсолютно доминирует над F_2 , если любая другая фонема F_1 , доминирующая над F_2 , отличается от F_2 большим количеством признаков, чем F_1 . Если F_1 абсолютно доминирует над F_2 , то мы будем говорить, что F_1 и F_2 входят в один фонемный ряд.

Например, *g* входит в один фонемный ряд с *k*. Здесь фонемный ряд совпадает с элементарной категорией и на первый взгляд мы не получили ничего нового.

Но этот же принцип можно, по-видимому, применить и к анализу русских гласных, для которых теория московской фонологической школы и приобретает подлинную ценность.

В самом деле, *o* абсолютно доминирует над *a*, и обе фонемы входят в один фонемный ряд, хотя они при нашем задании системы признаков и не входят в одну элементарную категорию, так как различаются более чем одним признаком.

Необходимо еще раз напомнить, что мы дали определение не для фонемного ряда в понимании Р. И. Аванесова, а построили некоторый формальный аналог этого понятия. Речь шла лишь о том, чтобы показать, что анализ чисто дистрибутивных черт, соответствующих некоторой реальной единице, вполне возможен.

§ 12. Парадигматическая модель

Приступим теперь к анализу отношений между фонемами, причем нас теперь будут интересовать не отношения в речи, а отношения в системе, т. е. парадигматические отношения. Детальное исследование парадигматики фонем было проведено в «Основах фонологии» Н. С. Трубецким. Мы ограничимся лишь некоторой простейшей ситуацией, более удобной с точки зрения моделирования.

Возьмем все фонемы данного языка, насчитывающие ровно *k* признаков каждая. Эти фонемы составляют некоторую фонологическую подсистему. В дальнейшем мы будем заниматься данной фиксированной подсистемой.

Примечание. В системе акустических признаков Р. Якобсона все фонемы, по-видимому, входят в одну подсистему.

Естественно желание как-то упорядочить фонемы внутри подсистемы. Обычно в фонологической литературе принято давать таблицы, в которых каждая фонема соответствует пересечению ряда признаков (в системе, предложенной Р. Якобсоном, таблицу заменяет так называемая матрица идентификации).

Например, подсистема шумных русского языка может быть представлена таблицей со следующими (несколько упрощенными) признаками:

		Губной		Переднеязычный		Среднеязычный		Заднеязычный	
		глухой	звонкий	глухой	звонкий	глухой	звонкий	глухой	звонкий
Смычный	твердый	<i>p</i>	<i>b</i>	<i>t</i>	<i>d</i>			<i>k</i>	<i>g</i>
	мягкий	<i>p'</i>	<i>b'</i>	<i>t'</i>	<i>d'</i>			<i>k'</i>	<i>g'</i>
Щелевой	твердый	<i>f</i>	<i>v</i>	<i>s</i>	<i>z</i>	<i>ʃ</i>	<i>ʒ</i>	<i>çh</i>	
	мягкий	<i>f'</i>	<i>v'</i>	<i>s'</i>	<i>z'</i>	<i>ʃ'</i>	<i>ʒ'</i>		
Аффрикативный	твердый			<i>c</i>		<i>ç</i>			
	мягкий								

Подобные таблицы очень удобны для практического использования (в частности, читателю рекомендуется в дальнейшем сравнивать предлагаемые ниже построения с этой таблицей). Однако подобные таблицы имеют тот недостаток, что по ним трудно составить себе представление о группировках фонем в некоторые классы и об отношениях между этими классами. Поэтому в дальнейшем предлагается несколько более сложный, но более удобный с указанной точки зрения способ упорядочения фонем. Он задается предписанием («алгоритмом»), состоящим из шести пунктов.

1. Выберем две фонемы, имеющие общую архифонему α_1 и присоединим к этому классу все другие фонемы, имеющие данную архифонему. Получим фонологический класс $K(\alpha_1) = K_1$. Заметим, что любой такой класс однозначно определяется произвольной фонемой x , входящей в него; поэтому мы иногда будем обозначать класс, в который входит фонема x , как $K(x)$.

П р и м е р. Возьмем из нашей подсистемы фонемы p и t . В обе фонемы входит в качестве архифонемы совокупность признаков «смычный», «глухой», «твердый». Различаются они признаками «губной» для p и «переднеязычный» для t . Присоединим сюда еще фонему k , имеющую ту же архифонему. Этими тремя фонемами исчерпывается класс фонем, имеющих данную архифонему. Как мы выше условились, этот класс может быть обозначен $K(p)$, или $K(t)$, или $K(k)$.

2. Пусть уже построено i классов. Выберем из оставшихся фонем максимальную совокупность фонем, таких, что: а) все они имеют общую архифонему α_{i+1} ; б) в каком-нибудь из ранее построенных классов для любой фонемы этой совокупности

найдется фонема, которая имеет с ней общую архифонему. Таким образом, мы построим K_{i+1} .

3. Оставшиеся фонемы, которые не поддаются нашему упорядочению, мы исключим из нашей подсистемы и проведем для них новое упорядочение.

Пр и м е р. Мы получили класс $K_1 = \{p, t, k\}$. По пункту 2 получим новые классы: $K_2 = \{b, d, g\}$, $K_3 = \{p', t', k'\}$, $K_4 = \{b', d', g'\}$, $K_5 = \{c, \acute{c}\}$, $K_6 = \{f, s, \acute{s}, ch\}$, $K_7 = \{f', s', \acute{s}'\}$, $K_8 = \{v, z, \acute{z}\}$, $K_9 = \{v', z', \acute{z}'\}$. Новое упорядочивание нужно, по видимому, провести для сонорных, которые входят, таким образом, в иную подсистему, и для j , который, видимо, сам образует собственную подсистему.

Итак, мы получили некоторое разбиение на классы в пределах подсистемы, причем каждый класс однозначно определяется своей архифонемой. Набор архифонем $\alpha_1, \alpha_2, \dots, \alpha_i$, полученный нашей процедурой упорядочивания, однозначно определяет некоторое разбиение на классы $K_1^{(1)}, K_2^{(1)}, \dots, K_i^{(1)}$, где верхний индекс указывает номер разбиения, а нижний индекс номер класса.

4. Выбрав некоторую архифонему β_1 , которая не совпадает ни с одной из взятых выше α_i , и повторив процедуру, предусмотренную в пунктах 1—2 (пункт 3 действителен только для исходного разбиения — о принципах выбора этого исходного разбиения см. несколько ниже), мы или исчерпаем все фонемы и получим новое разбиение на классы, каждый из которых определяется некоторой архифонемой β_i , или не получим разбиения на классы, поскольку не будут исчерпаны все фонемы подсистемы. Во втором случае мы будем считать каждую из числа оставшихся фонем изолированным классом и тем самым получим разбиение всей подсистемы на классы, которое определяется (с точностью до изолированных классов) набором архифонем $\beta_1, \beta_2, \dots, \beta_m$.

5. Теперь мы снова повторяем процедуру, описанную в пункте 4, и получаем новое разбиение на классы, которое определяется (с точностью до изолированных классов) набором архифонем $\gamma_1, \gamma_2, \dots, \gamma_k$.

6. Процедура, описанная в пункте 5, повторяется до тех пор, пока не будут исчерпаны все архифонемы.

Назовем теперь мощность класса количеством фонем в данном классе, а мощностью разбиения, мощность класса, наиболее часто встречающуюся в данном разбиении. Будем упорядочивать нашу систему разными способами.

Условимся в качестве исходного разбиения выбирать разбиение с наибольшей мощностью, в качестве второго — разбиение с оставшейся наибольшей мощностью и т. п.

$K^{(1)}$ — есть разбиение мощности три, и эта мощность наибольшая. Поэтому оно и выбрано в качестве исходного.

Построенная нами абстрактная система упорядочения фонем удобна для формальных построений, однако именно в силу ее абстрактности ее трудно соотносить с привычными фонологическими представлениями. Для многих читателей, по-видимому, может оказаться более удобным следующий принцип упорядочения. Возьмем какую-то архифонему α и максимальный набор однородных признаков m_1, m_2, \dots, m_t , таких, что α плюс любой из этих признаков составляют фонему данного языка. Теперь берем все фонемы, имеющие общую архифонему α_1 и различающиеся лишь признаками приведенного ряда. Так поступаем до тех пор, пока не исчерпаны все фонемы подсистемы, объединяемые в классы по совокупности переменных признаков. Мы будем говорить, что эта совокупность переменных признаков определяет некоторое разбиение, а архифонемы α, α_1 и т. д. определяют классы этого разбиения. Выбрав другую максимальную совокупность однородных признаков, мы получим новое разбиение и т. д.

Так, в нашем примере подсистемы шумных первое разбиение определяется совокупностью однородных признаков места образования, второе разбиение — признаками участия и неучастия голоса, третье разбиение — признаками твердости и мягкости и, наконец, четвертое — совокупностью однородных признаков способа образования.

Можно было бы показать, что данный принцип упорядочения логически следует из предыдущего. В связи с этим наша система допускает целый ряд интерпретаций, хорошо известных в фонологии.

Первая из них состоит в следующем. Каждое отдельное разбиение фонем можно представить как некоторое измерение (например, как прямую, точками которой являются отдельные классы данного разбиения). Каждая фонема представляется тогда как упорядоченная n -ка координат, где i -я координата указывает на класс, в который попадает данная фонема при i -м разбиении.

В нашем примере каждая фонема определяется четырьмя координатами:

$$p = (K_1^{(1)}, K_1^{(2)}, K_1^{(3)}, K_4^{(4)})$$

$$b = (K_2^{(1)}, K_1^{(2)}, K_2^{(3)}, K_5^{(4)}) \text{ и т. д.}$$

Если распространить такое представление на всю систему (добавив в качестве еще одной координаты номер подсистемы), то мы получим представление фонологической системы в качестве n -мерного пространства, близкое к рассмотренному

Черри¹¹. Мы могли бы даже получить представление, изоморфное представлению Черри, если бы исходили из упомянутых выше акустических признаков Якобсона. При этом, как мы указывали, наша подсистема охватывала бы всю систему. Кроме того, признаки Р. Якобсона выбраны так, чтобы в каждый класс любого разбиения попало бы не более двух фонем (как в разбиениях $K^{(2)}$ и $K^{(3)}$ нашего примера). Это позволяет говорить о признаке, принимающем ровно два значения: нуль и единицу. Подобная концепция, получившая название бинаризма, несомненно, привлекает своей строгостью, хотя и подвергается критике¹². Как указано, мы предпочитаем, однако, более традиционную интерпретацию признаков.

Такая, более традиционная, интерпретация существенна и потому, что она позволяет связать задание разбиений с классическим анализом фонологических оппозиций в смысле Н. С. Трубецкого. Так, если при заданном разбиении в каждый класс входит не более двух фонем, то мы имеем одномерную оппозицию, в противном случае мы имеем многомерную оппозицию. Дальнейший перевод терминологии этого параграфа на язык противопоставлений Н. С. Трубецкого легко может быть осуществлен самим читателем.

§ 13. Фонологическая однородность. Понятие пустых клеток

Перейдем к характеристике построенной системы. Ясно, что разные системы различаются разной степенью упорядоченности. С типологической точки зрения полезно иметь некоторую идеально упорядоченную систему, с которой можно сравнивать реальные языки.

О п р е д е л е н и е. Мы будем говорить, что некоторая подсистема является фонологически однородной, если для любых двух разбиений и для любых фонем x и y из условия

$$K^{(1)}(x) \cap K^{(2)}(y) \neq 0$$

(т. е. из наличия некоторой фонемы z , которая при разбиении $K^{(1)}$ принадлежит к тому же классу, что и x , и при разбиении $K^{(2)}$ принадлежит к тому же классу, что и y) следует

$$K^{(1)}(y) \cap K^{(2)}(x) \neq 0.$$

Интересно установить условия, при которых подсистема оказывается фонологически однородной. Наиболее важным

¹¹ См.: E. Colin Cherry. Roman Jakobson's «Distinctive features» as the normal coordinates of a language. «For Roman Jakobson». Haag, 1956.

¹² См.: А. Мартине. Принцип экономии в фонетических изменениях. Перевод с франц. М., 1960, стр. 102—103.

условием является отсутствие «пустых клеток». Мы будем говорить, что два признака m_i и m_j полностью однородны в данной подсистеме, если замена m_i на m_j (и m_j на m_i) в любой фонеме данной подсистемы приводит к фонеме данной подсистемы. Если в подсистеме есть две фонемы F_1 и F_2 , такие, что F_1 получена из F_2 заменой m_i на m_j , причем m_i и m_j не полностью однородны в данной подсистеме, то мы будем говорить, что подсистема содержит пустую клетку. При этом под пустой клеткой мы понимаем следующее образование: раз m_i и m_j не полностью однородны, то существует фонема F , такая, что замена m_i на m_j (или m_j на m_i) приводит к пустой клетке, а именно совокупности признаков, не составляющей фонемы данного языка.

Теорема 2.1¹³. Если подсистема не содержит пустых клеток, то она фонологически однородна.

Относительно рассмотренной подсистемы русских шумных мы не можем утверждать, что она однородна, ибо в ней сразу же устанавливаются пустые клетки. Но это еще не означает, что система фонологически неоднородна. Выяснить это наша теорема не помогает, и мы должны обращаться к анализу отдельных случаев. Но случаи неоднородности в ней действительно можно обнаружить.

Возьмем фонему t . По разбиению $K^{(4)}$ она относится к $K_1^{(4)} = K^{(4)}(c)$, а по разбиению $K^{(1)}$ относится к $K_1^{(1)} = K^{(1)}(p)$. Иначе говоря,

$$K^{(4)}(c) \cap K^{(1)}(p) = t.$$

В то же время нет фонемы, которая одновременно принадлежала бы к $K^{(4)}(p)$ и $K^{(1)}(c)$. Такая фонема, вообще говоря, возможна: это аффриката pf , как она произносится в немецком слове *Pfahl*. Для русского языка вместо фонемы pf мы имеем пустую клетку, связанную с тем, что замена признака «смычность» признаком «аффрикативность» не приводит к образованию фонемы, имеющейся в данном языке.

Мы видели, что теорема 2.1 позволяет лишь утверждать, что некоторая подсистема является фонологически однородной, но не позволяет отрицать однородность некоторой подсистемы. Существует, однако, более сильный признак однородности.

Теорема 2.2. Если подсистема фонологически однородна, то два регулярных класса одного разбиения не могут содержать разного количества фонем.

С л е д с т в и е. Фонологически однородная подсистема не содержит изолированных классов первого рода.

¹³ Первая цифра означает номер главы, вторая — порядковый номер теоремы в данной главе (доказательства этой и следующих теорем см. в приложении).

Теперь мы можем сразу сказать, что некоторая система не является фонологически однородной. Например, в нашем примере: $K_6^{(1)} = \{f, s, \check{s}, ch\}$, а $K_7^{(1)} = \{f', s', \check{s}'\}$, что достаточно для того, чтобы подсистема была фонологически неоднородной. Может оказаться полезным и следующий признак.

Теорема 2.3. Если система фонологически однородна и $K^{(1)}(x)$ есть изолированный класс при некотором исходном разбиении $K^{(1)}$ и y входит в тот же класс, что и x при исходном разбиении, то $K^{(1)}(y)$ также есть изолированный класс.

По этому признаку система нашего примера также не является фонологически однородной.

Так, $K_{10}^{(4)} = \{g\}$ есть изолированный класс, b входит в тот же класс, что и g при исходном разбиении, а именно $K_2^{(1)}$. В то же время $K^{(4)}(b)$ не есть изолированный класс. На этом примере видно и наличие пустой клетки, а именно отсутствие щелевой заднеязычной звонкой фонемы в русском языке.

Все предыдущие признаки фонологической однородности практически, по-видимому, не очень удобны, так как требуют предварительного упорядочения подсистемы фонем. Мы теперь сформулируем признак, который гораздо удобнее для применения, но для этого нам понадобится одно вспомогательное понятие.

Мы будем называть две фонемы соседними, если от одной из них можно перейти к другой, заменив ровно один признак. Количество фонем, соседних по отношению к данной фонеме, назовем рангом данной фонемы в данной подсистеме. Мы будем называть подсистему полной, если в ней любые две фонемы можно соединить цепочкой, в которой каждые две рядом стоящие фонемы соседние¹⁴.

Теорема 2.4. В фонологически однородной подсистеме две соседние фонемы имеют одинаковый ранг.

С л е д с т в и е. В полной фонологически однородной подсистеме все фонемы имеют одинаковый ранг.

Таким образом, для полной фонологически однородной подсистемы можно ввести понятие ранга подсистемы, а именно определить это понятие как величину, совпадающую с рангом произвольной фонемы.

Рассмотрим теперь один из предыдущих примеров. Фонема t имеет в нашей подсистеме ранг 6 (соседние фонемы: p, k, d, t', s, c), а фонема p имеет ранг 5 (соседние фонемы: t, k, b, p', f).

Этого оказывается достаточным, чтобы наша подсистема была фонологически неоднородной.

¹⁴ О важности принципа полноты, первые указания на который есть еще у Бодуэна де Куртенэ, см.: В. Н. Топооров. И. А. Бодуэн де Куртенэ и развитие фонологии. В сб. «И. А. Бодуэн де Куртенэ». М., 1960, стр. 35.

Интересно, что в чешском языке ранг фонемы t , как и в русском, равен 6 (соседние фонемы: p, k, t', d, s, c), а ранг фонемы p равен всего 4 (соседние фонемы t, k, b, f).

Это показывает, что понятие ранга может оказаться полезным при типологическом сравнении близкородственных языков.

Понятие полноты системы фонем позволяет сформулировать еще одно важное утверждение.

Теорема 2.5. Если подсистема является полной и фонологически однородной, то она не содержит пустых клеток.

Практически это утверждение дает нам (вместе с теоремой 2.1) критерий фонологической однородности, а именно отсутствие пустых клеток, так как обычно рассматриваемые в фонологии подсистемы полные. В частности, полной является подсистема из нашего примера.

Понятие однородности важно по следующим соображениям. В каждой нетривиальной подсистеме можно выделить подмножество фонем, образующих фонологически однородную подсистему. Максимальное подмножество таких фонем мы будем называть ядром фонологической подсистемы.

Так, в подсистеме нашего примера можно выделить ядро:

$$\begin{array}{lll}
 K_1^{(1)} = \{p, t, k\} & K_1^{(2)} = \{p, b\} & K_1^{(3)} = \{p, p'\} \\
 K_2^{(1)} = \{b, d, g\} & K_2^{(2)} = \{t, d\} & K_2^{(3)} = \{b, b'\} \\
 K_3^{(1)} = \{p', t', k'\} & K_3^{(2)} = \{k, g\} & K_3^{(3)} = \{t, t'\} \\
 K_4^{(1)} = \{b', d', g'\} & K_4^{(2)} = \{b', p'\} & K_4^{(3)} = \{k, k'\} \\
 & K_5^{(2)} = \{t', d'\} & K_5^{(3)} = \{d, d'\} \\
 & K_6^{(2)} = \{k', g'\} & K_6^{(3)} = \{g, g'\}
 \end{array}$$

Количество разбиений внутри подсистемы мы будем, имея в виду указанную интерпретацию нашей системы в качестве n -мерного пространства, называть размерностью подсистемы. Размерность подсистемы и размерность ядра, а также соотношения между ядром и всей подсистемой могут служить важными типологическими характеристиками языка.

Нам представляется, что предложенная терминология может послужить уточнением вопроса о функциональной нагрузке противопоставлений (в ядро как раз входят группы фонем, различающиеся наиболее полно нагруженными признаками).

Здесь мы приближаемся к идее создания модели, близкой к рассмотренной и формализующей подход диахронической фонологии. Эта идея (при несколько иной терминологии) принадлежит французскому лингвисту Мартине¹⁵.

¹⁵ А. Мартине. Указ. соч., глава III, «Структура».

§ 14. Динамическая модель Мартине

Мартине рассматривает два основных разбиения всего множества фонем, а именно разбиение на с е р и и и разбиение на р я д ы. Эти разбиения легко представить как интерпретацию наших абстрактных разбиений. Легко видеть, что разбиение К⁽¹⁾ из нашего русского примера соответствует разбиению на серии, а остальные наши разбиения соответствуют разбиению на ряды. В этом легко убедиться, если привести определения Мартине: «Мы будем говорить, что согласные фонемы, характеризующиеся одним и тем же признаком, образуют серию, если они артикулируются в различных точках, расположенных вдоль экспираторного канала... Мы будем говорить, что фонемы, которые артикулируются в одной и той же точке экспираторного канала, но различаются между собой каким-то иным характерным признаком, образуют ряд...»

Что касается гласных, то представляется удобным называть серией совокупность фонем, которые характеризуются одним и тем же типом резонирующей полости (например, так называемые передние гласные характеризуются очень малой резонирующей полостью в передней части рта), но различаются по степени открытости рта; „рядом“ же мы будем называть совокупность фонем, характеризующихся одной и той же степенью открытости рта, но различающихся по типу резонирующей полости¹⁶.

Термин «фонологическая однородность» Мартине не употребляет, но его идея целиком соответствует этому понятию, тем более что он пользуется понятием «пустой клетки» в том же значении.

Мы попытаемся наметить путь к формализации идей Мартине в некоторой кибернетической модели.

Пусть количество пустых клеток характеризует степень неустойчивости, лабильности системы и пусть подсистема имеет ряд состояний равновесия, которые характеризуются фонологической однородностью и могут быть достигнуты как исключением некоторых более или менее изолированных фонем, так и созданием новых фонем (за счет того, что связанные признаки становятся релевантными и варианты превращаются в новые фонемы). Пусть, с другой стороны, подсистема, стремящаяся достичь фонологической однородности, находится под влиянием «возмущений» со стороны, которые приводят к образованию новых фонем и новых пустых клеток. Тогда каждый период в развитии языка характеризуется некоторой идеальной моделью, а именно фонологически однородной системой, достигаемой самым близким путем (т. е. наименьшим числом операций).

Возьмем в качестве примера действия такой модели под-

¹⁶ Там же, стр. 47.

систему гласных (без дифтонгов) современного немецкого языка, как она описывается действующими орфоэпическими нормами¹⁷.

Выделим следующие совокупности по $n - 2$ признаков (архифонемы «второго порядка»):

α_1	общие признаки для	$i:$, i
α_2	»	»
α_3	»	»
α_4	»	»
α_5	»	»
α_6	»	»
α_7	»	»

Эти архифонемы выбраны потому, что внутри них признаки сочетаются по некоторым законам, не зависящим от данного конкретного языка (так лабиализованными могут быть лишь гласные переднего ряда). В модели мы от таких связей между признаками вообще отвлекались. Остаются две пары признаков: открытость — закрытость и долгота — краткость. Если бы внутри каждой пары признаки были полностью однородны, мы могли бы из 7 архифонем второго порядка получить 28 гласных фонем. На самом деле мы имеем 15 фонем. Количество пустых клеток очень велико. Однако если мы выделим подсистему из 14 фонем, исключив, например, $\epsilon:$, то мы получим (в предположении об однородности архифонем) некоторое фонологически однородное ядро, в котором, правда, один из признаков, например, долгота — краткость, избыточен. Интересно, что подобный процесс и происходит в современном немецком разговорном языке, где $\epsilon:$ и ϵ : начинают совпадать¹⁸.

Близость к фонологической однородности может быть достигнута не только путем изменения количества фонем, но и путем переопределения заданной системы признаков. Рассмотрим, например, в чешском языке подсистему шумных: b , p , v , f , s , z , d , t , g , ch , k , \check{s} , \check{z} , \check{c} , c , t' , d' — всего 17 фонем. Если считать, что t и t' (соответственно d и d') различаются по признаку твердость — мягкость, то остальные 13 фонем будут непарными по этому признаку, и возникает 13 пустых клеток. Если мы, следуя ряду крупных фонологов, будем считать что t (и d) отличается от t' (и d') только по признаку места образования: t (d) содержит признак «зубной», а t' (d') признак «переднеязычный», то, как легко проверить, количество пустых клеток рез-

¹⁷ См.: Siebs. Deutsche Hochsprache. Bühnenaussprache. Hrsg. von H. de Boor und P. Diels. Berlin, 1958.

¹⁸ См.: W. Neumann. Sprachwandel und Sprachpflege. «Sprachpflege», 1961, Hf. 4, стр. 71.

ко сократится. По-видимому, этот аргумент должен учитываться при характеристике системы согласных чешского языка. Интересно с точки зрения нашей модели рассмотреть и спор о месте фонем \acute{c} , $d\acute{z}$, а также \acute{s} , \acute{z} в системе шумных современного польского языка. По своим фонетическим особенностям первые два звука — мягкие аффрикаты, а вторые — мягкие шипящие. Между тем в теоретической литературе¹⁹ часто указывается, что первые две фонемы надо рассматривать как t' , d' , а вторые как s' , z' . Оказалось, что при упорядочении польских фонем, признаки которых заданы строго фонетически, получается более чем вдвое больше пустых клеток, чем при перепределении признаков по принципу, указанному выше. Наиболее интересно, однако, рассмотреть с точки зрения нашей модели идею, высказанную недавно Штибером²⁰. Согласно этой идее \acute{s} (и соответственно \acute{z}) в современном польском языке надо считать одновременно мягким коррелятом к \acute{s} (соответственно \acute{z}) и мягким коррелятом к s (соответственно z). В терминах нашей модели эту идею можно выразить так, что максимальное приближение к однородности, т. е. сокращение числа пустых клеток иногда может быть достигнуто и путем расщепления одной фонемы на две (разумеется, такое расщепление должно быть подтверждено анализом вне модели).

Эти примеры показывают, что динамическая модель может рассматриваться как порождающая парадигматическая модель (по некоторым заданным фонемам порождаются новые).

В дальнейшем мы увидим, что абстрактная модель, близкая к рассмотренной здесь фонологической, может быть предложена и для грамматики (см. § 30).

§ 15. Синтагматические модели

В дальнейшем речь будет идти об отношении между фонемами, устанавливаемом на синтагматической оси в тексте. Великолепный образец такого функционального анализа был дан еще Н. С. Трубецким на примере старославянского языка²¹. Например, Трубецкой предложил классифицировать все согласные этого языка в зависимости от того, какие гласные могут комбинироваться с ними. Так, согласный N противопоставлялся всем остальным согласным, поскольку перед ним встре-

¹⁹ См., например: С. К. Шаумян. История системы дифференциальных элементов в польском языке. М., 1958, стр. 75—76; Л. Э. Калыныч. Развитие корреляции твердых и мягких согласных фонем в славянских языках. М., 1961, стр. 64, 67—68.

²⁰ См.: Z. Stieber. Über das gegenseitige Verhältnis der heutigen polnischen Phoneme $s - \acute{s}$ und $z - \acute{z}$. «Die Welt der Slaven», 1961, 2, стр. 121—124.

²¹ Н. С. Трубецкой. Altkirchenslavische Grammatik, Schrift-, Laut- und Formensystem. Wien, 1954, стр. 78.

чались лишь *o, ö, e, a*. Другие согласные распадались на три больших класса в зависимости от того, допускались ли перед ними только задние гласные, или только передние, или же и те и другие. В последней группе опять-таки на основе распределения выделялись две подгруппы и т. д.

Такой же метод анализа был продемонстрирован им на примере древнегреческого языка²². В американской дескриптивной лингвистике метод синтагматического анализа текста наиболее полно разработан в работе Блока и Трейджера²³, где введено важное понятие структурного объединения фонем (structural set), как класса фонем, встречающихся в одинаковом окружении.

Мы рассмотрим теперь модель, которая построена, исходя из наиболее простой ситуации, а именно: считается, что текст разбит всеми возможными способами на пары фонем, встречающихся в тексте рядом. Было замечено, что к полученному таким образом множеству упорядоченных пар фонем может быть приложен аппарат теории отношений²⁴.

В таблице приведены в виде матрицы пары фонем, встречающиеся в начале слова в четырех славянских языках. Крестик на пересечении некоторой строки и некоторого столбца обозначает, что в данном языке встречается пара, в которой на первом месте стоит фонема данного столбца и на втором месте фонема данной строки.

Вводятся, например, следующие понятия.

Степень полноты. Обозначим множество всех пар для данного языка через R . Под степенью полноты (D) мы будем понимать отношение числа пар, входящих в R (обозначим его через $n(R)$), к общему числу возможных пар k^2 , где k — число фонем. Тогда

$$D = \frac{n(R)}{k^2}.$$

Для нашей таблицы эта величина равняется:

Русский	Сербский	Польский	Чешский
$\frac{216}{35^2} \approx 0,18$	$\frac{104}{25^2} \approx 0,17$	$\frac{213}{36^2} \approx 0,16$	$\frac{221}{27^2} \approx 0,30$

²² См.: Н. С. Трубецкой. Основы фонологии, стр. 272—273.

²³ Bloch and Trager. Outline of linguistic analysis. Baltimore, 1942.

²⁴ См.: Fr. Harny and H. Papert. Toward a general calculus of phonemic distribution. «Language», vol. 33, 1957, № 2; B. Sigrud. Rank order of consonants established by distributional criteria. «Studia linguistica», 9, 1955, № 1.; J. Garding. Relations and order. Там же. E. M. Vasiliu. Une classification des consonnes roumaines d'après le critère de la distribution. «Mélanges linguistiques». Bucarest, 1957. Мы будем в дальнейшем в основном опираться на первую работу; об общей теории отношений см.: А. Тарский. Введение в логику и методологию дедуктивных наук. Перевод с англ. М., 1948, стр. 129 и сл.

Здесь интересно то, что первые три языка при разном инвентаре и разном числе фонем дают почти одинаковую степень полноты, в то время как чешский язык дает качественно иной показатель.

Конечно, какое-либо истолкование этих данных было бы преждевременным, главным образом потому, что исходная таблица недостаточно надежна и распространяется лишь на консонантные начала. Тем не менее в какой-то мере они отражают действительное положение в языке. Введение сочетаний с гласными вряд ли намного изменит картину, так как гласные в славянских языках довольно свободно сочетаются со всеми согласными.

Степень рефлексивности. Выделим во множестве пар фонем подмножество вида $\langle x, x \rangle$, т. е. такое, в котором на первом и втором местах стоят одинаковые фонемы (на таблице таким парам соответствуют крестики на большой диагонали). Для славянских языков соответствующее подмножество очень мало. В него входят:

Русский	Сербский	Польский	Чешский
$\langle s, s \rangle$	0	$\langle v, v \rangle$	$\langle s, s \rangle \langle z, z \rangle$

Все эти пары образованы за счет продуктивных славянских приставок и не играют существенной роли. Если мы введем степень рефлексивности как отношение числа рефлексивных пар $\langle x, x \rangle$ к общему числу фонем

$$P = \frac{n \langle x, x \rangle}{k},$$

то для славянских языков эта величина несущественна. Она может понадобиться для сравнения славянских языков с теми языками, где этот показатель велик.

Таково, по-видимому, положение в эстонском языке, где имеет смысл рассматривать «сверхдолгие» гласные и согласные как пары фонем²⁵ и где поэтому степень рефлексивности очень высока.

Было предложено²⁶ аналогичным образом исследовать подмножества пар, обладающих свойствами симметрии и транзитивности. Однако мы не будем здесь останавливаться на этих вопросах.

Задание пар фонем, встречающихся в данном языке, дает возможность построить ориентированный граф, отражающий распределение фонем в данном языке. А именно, каждая фонема представляет некоторую точку на плоскости.

²⁵ Ср. соображения по этому вопросу в диссертации: R. T. Harms. A descriptive grammar of Estonian. Chicago, Illinois, 1960, стр. 38—39.

²⁶ Fr. Nagary and H. Parer. Указ. соч.

	b	b'	c	č	d	d'	f	f'	g	g'	ch	ch'	j	k	k'	l	l'
b					+	+										+	+
b'																	
c																	
č											+						+
d																+	+
d'													+			+	+
f																+	+
f'													+				
g						+										+	+
g'																	
ch																+	+
ch'																	
i																	
k																+	+
k'																	
l	+	+							+								
l'						+			+								
m			+	+					+		+	+				+	+
m'																	
n																	
n'																	
p				+			+				+					+	+
p'																	
r			+		+	+											
r'		+											+				
s	+	+	+	+	+		+	+	+		+	+	+	+	+	+	+
s'											+	+					
š				+										+	+	+	+
t														+	+	+	+
t'																	
v	+		+	+	+				+	+	+			+	+	+	+
v'													+				
z	+	+			+	+			+	+						+	+
z'																	
ž	+				+				+							+	
Пауза	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

²⁷ W. Appel. Energiebasis — Artikulationsbasis. «Wiener Slavistisches

слова (по Апелю)

язык

т	т'	п	п'	р	р'	л	л'	с	с'	ш	т	т'	в	в'	з	з'	ж	Глас- ный
						+	+									+		+
		+												+				+
+							+				+	+	+					+
	+	+				+	+						+	+		+	+	+
			+															+
						+												+
	+	+	+			+	+						+		+			+
+	+	+				+	+						+					+
		+	+			+	+	+	+		+	+	+	+				+
									+	+					+		+	+
		+	+			+	+	+		+					+			+
						+												+
		+	+			+	+	+	+	+	+	+						+
											+		+	+			+	+
						+												+
		+	+			+	+	+	+	+	+	+						+
+	+	+	+	+	+	+	+	+			+	+	+	+	+		+	+
	+		+				+					+	+	+				+
+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+
+							+											+
	+	+	+			+	+							+				+
							+											+
			+			+	+							+				+
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

	<i>b</i>	<i>c</i>	<i>č</i>	<i>č̣</i>	<i>d</i>	<i>đ</i>	<i>đ̣</i>	<i>f</i>	<i>g</i>	<i>ch</i>	<i>i</i>	<i>k</i>	<i>l</i>	<i>l'</i>
<i>b</i>					+								+	+
<i>c</i>												+		
<i>č</i>														
<i>č̣</i>												+	+	
<i>d</i>	+												+	
<i>đ</i>	+													
<i>đ̣</i>														
<i>f</i>													+	
<i>g</i>					+								+	+
<i>ch</i>													+	+
<i>i</i>														
<i>k</i>			+										+	+
<i>l</i>														
<i>l'</i>														
<i>m</i>													+	+
<i>n</i>														
<i>n'</i>														
<i>p</i>				+									+	+
<i>r</i>														
<i>s</i>		+								+	+	+	+	+
<i>š</i>		+	+	+								+	+	+
<i>t</i>												+	+	
<i>v</i>					+								+	
<i>z</i>	+				+				+		+		+	
<i>ž</i>	+				+		+		+				+	+
Паяа	+	+	+	+	+	+	+	+	+	+	+	+	+	+

<i>m</i>	<i>n</i>	<i>n'</i>	<i>p</i>	<i>r</i>	<i>s</i>	<i>š</i>	<i>t</i>	<i>v</i>	<i>z</i>	<i>ž</i>	Глас- ный
				+							+
+				+				+			+
											+
+			+	+				+			+
+	+			+				+			+
											+
											+
				+							+
+	+	+		+				+			+
+				+			+	+			+
											+
+	+	+		+				+			+
											+
	+	+		+							+
											+
				+	+		+				+
											+
+	+		+	+			+	+			+
+		+	+				+	+			+
+				+				+			+
				+							+
+	+			+				+			+
+		+		+				+			+
+	+	+	+	+	+	+	+	+	+	+	+

	b	b'	c	ć	č	d	ǰ	ź	ż	f	f'	g	g'	h	ch	i	k	k'
b																		
b'																		
c																	+	
ć																		+
č				+														+
d	+																+	
ǰ	+																	
ź																		
ż												+						
f																		
f'																		
g	+					+	+											
g'																		
h																		
ch				+														
i																		
k				+														
k'																		
l												+						
ł	+											+					+	
m						+						+						+
m'																		
n																		
n'																		
p			+		+											+		
p'																		
r						+												
s			+		+					+						+		+
s				+														
ś					+													+
ś'																		+
t																+		+
v		+	+	+	+	+		+				+			+	+	+	+
v'																		
z	+	+	+		+	+		+	+			+	+	+		+		
z'								+										
ż	+					+			+			+						
Пауза	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

l	l'	m	m'	n	n'	p	p'	r	s	s'	š	t	v	v'	z	z'	ž	Глас- ный
+	+							+							+	+	+	+
+	+	+		+														+
		+											+	+				+
+		+										+	+					+
+	+	+		+	+			+					+					+
													+					+
+								+							+			+
																		+
+	+	+	+	+	+			+					+	+	+	+	+	+
																		+
								+				+						+
+	+	+						+				+	+		~		+	+
+	+	+		+	+			+	+	+	+	+	+				+	+
																		+
					+						+				+		+	+
+	+			+	+			+			+	+			+		+	+
																		+
																		+
+	+				+			+	+	+	+	+	+				+	+
																		+
																		+
+	+				+													+
																		+
																		+
+	+	+		+	+			+	+				+	+	+		+	+
+																		+
+	+	+		+	+													+
+	+	+	+	+	+													+
+	+	+	+	+	+	+	+		+	+	+	+	+		+	+	+	+
+	+	+		+	+			+	+									+
+	+	+	+	+	+													+
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

	b	c	č	d'	d	ž	f	g	h	ch	j	k	l
b				+									+
c													+
č												+	+
d'													
d	+	+											+
ž	+												
f													+
g					+								+
h	+												+
ch		+											+
j					+				+				
k					+								+
l	+								+			+	
m					+				+	+		+	+
\tilde{n}													
n													
p										+	+		+
r		+	+	+	+								
ř	+											+	
s	+	+	+	+	+		+	+	+	+	+	+	+
š			+									+	+
t'													
t			+							+		+	+
v	+	+	+	+	+				+	+	+	+	+
z	+	+	+	+	+		+	+	+	+	+	+	+
ž	+			+	+			+	+				+
Пауза	+	+	+	+	+	+	+	+	+	+	+	+	+

<i>m</i>	<i>ñ</i>	<i>n</i>	<i>p</i>	<i>r</i>	<i>ř</i>	<i>s</i>	<i>š</i>	<i>l'</i>	<i>l</i>	<i>v</i>	<i>z</i>	<i>ž</i>	Глас- ный
				+	+						+	+	+
+		+	+						+	+			+
+	+		+	+				+	+				+
													+
+	+	+		+	+					+			+
	+			+									+
	+	+		+									+
+	+	+		+	+					+			+
+	+	+		+	+			+		+			+
+						+							+
+	+	+		+	+	+	+		+	+			+
	+	+	+			+				+	+	+	+
	+	+		+	+	+	+				+	+	+
													+
													+
	+	+		+	+	+	+		+				+
+								+	+	+	+	+	+
										+			+
+	+	+	+	+	+	+	+		+	+	+	+	+
+	+	+	+	+				+	+	+			+
													+
+	+	+		+	+		+			+			+
+	+	+	+	+	+	+	+	+	+		+	+	+
+	+	+	+	+	+	+	+	+	+	+	+	+	+
+	+	+		+						+			+
+	+	+	+	+	+	+	+	+	+	+	+	+	+

Если в списке пар встретилось $\langle x, y \rangle$, то от x к y проводится стрелка.

Ниже будет (с некоторыми упрощениями) рассмотрена модель, исследованная румынскими учеными: математиком С. Маркусом и лингвистом Эм. Василиу²⁸. Пусть нам даны все фонетические слова данного языка в смысле интерпретации 1. Выразим слово в виде цепочки фонем. Перенумеруем все фонемы в каждом слове, начиная с конца, например:

fstr eča

7654321

Заметим, что две одинаковые фонемы, стоящие на разных местах, будут нами теперь рассматриваться как разные элементы.

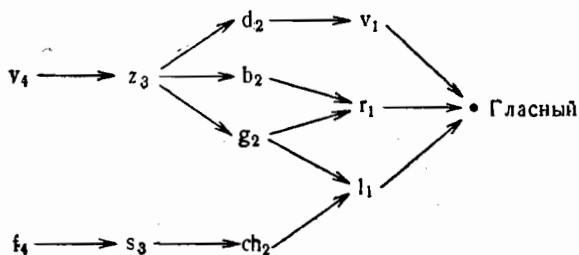
Возьмем теперь некоторую фонему f с индексом i . Выделим все пары вида f_i, f_{i-1} , где f_i — зафиксированная нами фонема, а f_{i-1} — фонема, которая хотя бы в одном фонетическом слове стоит после нее.

Проделав эту операцию для всего множества фонем, мы и построим искомый граф.

Для того чтобы дальнейшее изложение было достаточно конкретным и лингвистически значимым, мы ограничимся случаем начальных консонантных групп русского языка, причем возьмем только консонантные группы, начинающиеся с v или f .

Имеются, например, следующие группы, по четыре согласных: $v_4z_3b_2r_1$, $v_4z_3g_2l_1$, $v_4z_3d_2v_1$, $v_4z_3g_2r_1$ и $f_4s_3ch_2l_1$ (на письме *всхл*, например *всхлопнуть*).

Получаем следующий подграф



Подобная модель может, так же как и анализ пустых клеток в парадигматике, оказаться полезной для типологической классификации языков. Сравнивая между собой фонемы, представ-

²⁸ S. Marcus et E. m. Vasiliu. La théorie des graphes et le consonantisme de la langue roumaine. 1. «Académie de la république populaire roumaine. Revue de mathématiques pures et appliquées», t. V, 1960, № 2, стр. 319—340.

ленные близкими по звучанию звуками в двух разных языках, например русском и польском, можно подсчитать порядок соответствующего узла, а именно количество входящих и количество исходящих стрелочек.

Можно использовать для этого (отвлекаясь от разницы между двумя графами) данные таблицы: число входящих к данному узлу стрелок (Хэрари и Пейпер в указанной работе называют это «бета-полем фонемы x ») соответствует числу крестиков в данном столбце, а число исходящих стрелок («альфа-поле фонемы x ») — числу крестиков в данной строке. Ниже приводятся данные для более или менее близких по произношению звуков: z , s , \check{s} в четырех славянских, а также в румынском языке (по данным Маркуса и Василиу)

Звук	Язык	Количество входящих стрелок	Количество исходящих стрелок
z	русский	5	13
	сербский	0	7
	польский	6	22
	чешский	7	26
	румынский	0	8
s	русский	5	31
	сербский	2	13
	польский	5	14
	чешский	8	24
	румынский	0	8
\check{s}	русский	5	19
	сербский	0	11
	польский	6	11
	чешский	7	11
	румынский	0	8

Конечно, и этот материал служит лишь для иллюстрации и не дает права делать какие-нибудь выводы.

В связи с этой моделью поставлен также следующий интересный вопрос. В каких случаях, зная общее число фонем в данной цепочке, а также некоторые из фонем, можно восстановить остальные?

Например, из нашего чертежа ясно, что если нам даны две фонемы: v_4 и v_1 , то две промежуточные легко можно восстановить; аналогично в случае, если даны v_4 и l_1 . Имеется и более интересный случай: если дано f_4 , то все три следующие

фонемы можно восстановить. Если же нам даны v_4 и r_1 , то мы можем восстановить z_3 , но не можем восстановить следующую за ней фонему. Из чертежа ясно, почему это происходит. В случае, когда по некоторым двум фонемам можно восстановить промежуточные, от одной фонемы к другой ведет один путь и обратно, когда имеется один путь, то по двум фонемам можно восстановить промежуточные ²⁹.

§ 16. Модель порождения

До сих пор мы рассматривали фонетические слова (в смысле обеих интерпретаций § 9) как исходный материал и, основываясь на них, строили в дальнейшем всю теорию, т. е. все рассмотренные выше модели были анализирующими. Посмотрим теперь, как может быть построена синтезирующая модель. Простейшим способом построения такой модели было бы обращение предыдущей модели в следующем смысле.

Пусть у нас имеется некоторый граф (например, построенный на основании данных таблицы).

Будем теперь двигаться от точки, обозначающей пробел (паузу) по стрелкам графа до некоторой заключительной точки (если бы наша таблица была дополнена сочетаниями с гласными и согласными в середине и конце слова, то такой заключительной точкой была бы новая пауза). Выпишем все фонемы, которые мы прошли, двигаясь по графу. Получим некоторую последовательность фонем. Двигаясь другим путем, мы получим другую последовательность и т. п. В каком отношении стоят эти цепочки к словам соответствующего языка?

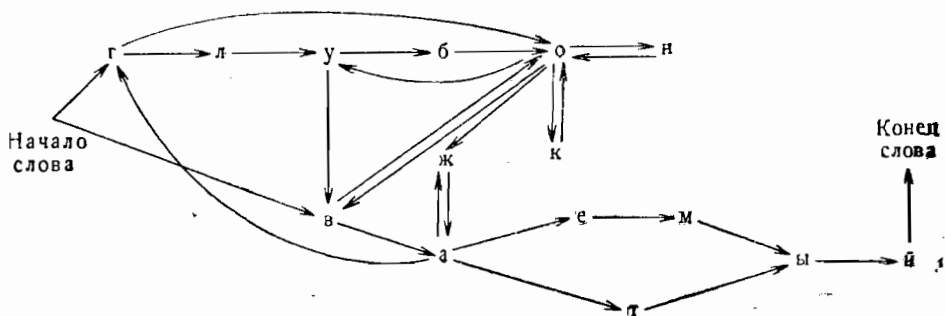
Во-первых, ясно, что работа данного устройства приводит не к синтезу фонетических слов, а к синтезу цепочек фонем, т. е. абстрактных наборов релевантных признаков. Чтобы перейти от фонем к звукам речи, из которых состоят фонетические слова, мы должны восстановить и все нерелевантные признаки. Тогда мы можем считать, что наша модель действительно синтезирует некоторые фонетические слова.

Во-вторых, здесь наблюдается интересное несоответствие между множеством анализируемых фонетических слов и множеством синтезируемых фонетических слов. Мы условились, что анализироваться будут реальные слова данного языка (интерпретация 1). Иначе говоря, наш граф отражал сочетания, которые встретились в реальных словах языка. Между тем син-

²⁹ Маркус и Василиу называют в указанной работе фонемы, восстанавливаемые таким путем, и з б ы т о ч н ы м и. Действительно, имеется некоторая связь между проблемой избыточности, как она ставится в современной теории информации, и рассматриваемой нами. Однако мы предпочитали бы не употреблять в этой связи термина «избыточность».

тезирующее устройство, действующее описанным нами способом, выдает не только то, что реально встретилось в языке, но и то, что было потенциально заложено в нем.

Например, проанализировав слова *вагоновожатый* и *глубокоуважаемый* (при условии, что фонемы не нумеруются, а это вполне допустимо, поскольку мы нумеровали их лишь внутри сочетаний согласных, которых здесь почти нет), мы получим (в буквенном представлении) следующий граф:



Если теперь реализовать устройство, описанное выше, то оно произведет не только знаменитое *вагоновожатый* *глубокоуважаемый*, но и целый ряд других слов, таких как *глухоемкий*, *глубокоемкий* и т. п.

Модель порождает не встретившиеся в реальном тексте слова. Если учитывать не только возможные пары двух элементов (здесь мы для простоты возьмем буквы, а не звуки) в русском языке, но и условные вероятности появления некоторой буквы после данной, то, как показал Р. Л. Добрушин, можно прийти к фразе типа *Умарано кач всванный рося ных ковкров недаре*. Даже если учитывать вероятность появления трехбуквенных сочетаний, то получаются все еще бессмысленные сочетания типа *Покак пот дурноскака наконецно зне ствоовил се твоей обниль*. Только учет вероятности появления четырехбуквенных сочетаний приводит к фразам, близким к «зауми», о которой говорилось в § 9: *Весел вратья не сухом и непо и корко*³⁰.

Заметим теперь, что положение не изменилось бы существенно, если бы вместо букв стояли звуки речи³¹. Между тем в четырехэлементных сочетаниях уже вполне четко проявля-

³⁰ Результаты опытов Р. Л. Добрушина приведены по кн.: А. М. Яглом и И. М. Яглом. Вероятность и информация. 2-е изд. М., 1960, стр. 192 и сл. (см. также: «Математическое просвещение», вып. 6. М., 1961, стр. 48—49).

³¹ Для сочетаний из двух звуков соответствующие вероятности подсчитаны Р. Л. Зиндером, см. «Вопросы статистики речи», Л., 1957, стр. 58—61.

ются все позиционные особенности, характерные именно для данного языка. С другой стороны, даже используя вероятности появления многоэлементных (10 и более) сочетаний, мы не обязательно получим действительно зафиксированные в языке фонетические слова.

Тем самым подтверждается полезность интерпретации исходного фонетического материала как множества потенциальных фонетических слов.

В заключение отметим, что моделью, аналогичной рассмотренной выше, является модель языка с конечным числом состояний, которая исследуется в § 38.

ОСНОВНЫЕ ПОНЯТИЯ ПРИ ПОСТРОЕНИИ
ГРАММАТИЧЕСКИХ МОДЕЛЕЙ
И ИХ ЛИНГВИСТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ

§ 17. Морфема или слово?

В качестве интерпретации понятия «элемент» в фонологической модели мы легко приняли понятие звука речи. В грамматике дело обстоит не так просто. При интерпретации формального понятия «элемент» можно иметь в виду две разные единицы, а именно морфему и слово. Интерпретация будет существенно меняться в зависимости от выбора этого исходного понятия. По-видимому, от этого же зависит и выбор формально более удобного аппарата.

Обратимся поэтому к вопросу о том, какая единица более удобна для формального анализа прежде всего с точки зрения критерия выделяемости в потоке речи.

В исследованиях последнего времени, как европейских, так и американских, в качестве исходной единицы берется, обычно без строгого определения этого понятия, морфема (в смысле «минимальная значимая единица», иногда в этом значении употребляется термин «монема»¹, а термин «морфема» обозначает лишь грамматические или формальные монемы). При этом очевидно, что морфема есть единица языка, а в речи выступают отдельные представители морфем или алломорфы, которые удобно называть «морфами». Например морфы *vod*, *vot*, *vad* (ср. *воды*, *вод*, *вода*) являются алломорфами одной морфемы как единицы языка.

Разработан ряд формальных процедур выделения морфов (на некоторых из них мы остановимся в дальнейшем). Однако все эти процедуры еще слишком сложны, а главное, недостаточно точны для того, чтобы понятие морфемы, далеко не столь очевидное, как понятие слова, принимать в качестве исходного.

¹ H. F r e i. Critères de délimitation. «Word», vol. 10, 1954, № 2—3.

Выбор морфемы в качестве исходной единицы осложняется следующим обстоятельством. Как известно, морфы по-разному связаны внутри слов или речевых тактов. Как мы покажем ниже, во многих языках одни группы морфем (окончания, отчасти суффиксы) теснее связаны с основой, чем другие (префиксы, компоненты сложных слов). В связи с этим выделение морфем какой-нибудь эффективной процедурой без обращения к смыслу крайне затруднено.

Между тем с е г м е н т а ц и ю, членение речевого потока следует проводить на основании ясных формальных критериев. Мы собираемся теперь показать, что некоторые единицы, для ряда языков близкие к слову (в смысле «словоформа»), но, вообще говоря, занимающие промежуточное положение между морфем и словом, могут быть выделены на основе двух достаточно простых формальных критериев: фонологического и конструктивного.

§ 18. Фонологические критерии сегментации

Как показал Н. С. Трубецкой², имеется ряд фонологических сигналов, которые обозначают границу слов или морфем. Эти пограничные сигналы (*Grenzsignale*) могут быть единичными или групповыми. Примерами единичных сигналов являются твердый приступ в немецком языке, в южных диалектах польского языка, в некоторых диалектах чешского языка, в армянском и других языках, фиксированное ударение в немецком, чешском, армянском, польском и других языках, появление специальных позиционных вариантов фонем, например аспирированных *p'*, *t'*, *k'* в корневых морфемах немецкого языка. Групповыми сигналами являются специальные сочетания звуков, например в немецком или эстонском языке сочетание согласный + *h* (другие примеры см. в указанной работе Н. С. Трубецкого).

Важнейшим пограничным сигналом являются также паузы.

Будем теперь считать, что для исследуемого языка нам списком заданы все пограничные сигналы или хотя бы важнейшие из них. Конечно, некоторые из перечисленных выше сигналов были получены на основании знания о границах слов. Однако, во-первых, мы могли бы ограничиться лишь теми, для получения которых никаких предварительных знаний о членении текста не требуется (например, паузы и сведения, полученные на основе анализа распределения пауз). Во-вторых (и это, по-видимому, самое главное), задание списком фонологических пограничных сигналов вовсе не нарушает

² Н. С. Трубецкой. Основы фонологии. Перевод с нем. М., 1960.

строгости изложения, ибо не приводит к каким-либо кривотолкам в дальнейшем.

Ясно, однако, что одним фонологическим критерием обойтись нельзя, ибо во многих языках, например французском, границы между словами или морфемами вообще, как правило, не обозначаются, в других же языках эти границы обозначены далеко не во всех случаях.

Поэтому необходимо ввести в рассмотрение еще один критерий сегментации.

§ 19. Конструктивный критерий сегментации

Мы будем говорить, что между двумя фонемами *x* и *y* прохода, если можно опустить часть фразы, включающую *x*, или часть фразы, включающую *y*, без изменения грамматической правильности фразы. Например, во фразе *он сегодня перестарался* можно провести следующие границы: *он/сегодня/перестарался*. В немецкой фразе *der Täter verachtet diese Verordnung* границы проходят следующим образом: *der/Täter/ver/achtet/diese/Ver/ordnung* (ср. в русском *преступник/презирает/это/приказание*).

В разных языках эти критерии по-разному связаны между собой. Но обычно — и это один из поразительных лингвистических феноменов, еще требующий объяснения — фонологический критерий подкрепляется конструктивным, и наоборот. Приведем примеры.

В русском и других славянских языках наличие границы (по нашему конструктивному критерию) внутри слова явление весьма редкое; приставки обычно тесно спаяны с основой, ср. хотя бы частые явления парасинтетизма типа *наплясаться, побережье, подоконник, предсердье*. Одновременно группой фонем, спаянной целым рядом фонетических признаков, единицей, в которой, по меткому выражению А. А. Реформатского, разыгрываются основные законы позиционных изменений, является слово или даже более объемное образование, а именно речевой такт³. Еще более тесно спаяны между собой части слова в романских языках.

С другой стороны, в немецком и других германских языках наличие границ внутри слова явление весьма частое⁴. Достаточно напомнить о сложных словах немецкого языка типа *Arbeitsamt* 'биржа; труда', *Welthandel* 'мировая торговля' и т. п.,

³ См., например: А. Н. Гвоздев. О пределах действия звуковых закономерностей в русском языке. «О фонологических средствах русского языка». М., 1949, стр. 84—97.

⁴ См. в особенности: Ш. Балли. Общая лингвистика и вопросы французского языка. Перевод с франц. М., 1955, стр. 352—363, где последовательно сравниваются факты немецкого и французского языков.

в которых по конструктивному критерию легко определяется граница.

Эта граница очень часто совпадает с фонологической. Так, в наших примерах на границе появляется твердый приступ, например *fer'axtən, fer'ɔrdnuŋ 'arbaets'amt*, аспирированное *t*, ср. *t'aetər*, сочетание согласный +*h*, например *velthandəl*. Иногда между элементами сложного слова появляется столь важный фонологический критерий границы, как пауза, ср. у Шиллера в «Валленштейне»:

Jetzt muß
Gehandelt werden, eh die Glücks-
Gestalt mir wieder wegfiegt überm Haupt.

‘Теперь надо действовать, пока образ счастья вновь не улетит от меня’ — сложное слово *Glücksgestalt* оказалось разделенным из-за явления enjambement.

В немецком языке наблюдается, таким образом, резкое различие между окончаниями и суффиксами, с одной стороны, и компонентами сложных слов и префиксами, с другой стороны, первые не отделяются ни по фонологическому, ни по конструктивному критерию, в то время как вторые легко отделимы. В этом отношении немецкий язык не представляет какого-то исключения. Аналогичное положение во многих финно-угорских языках, и, в особенности, в финском или венгерском, где действие гармонии гласных, как правило, распространяется на окончание, но не распространяется на другой компонент сложного слова; например, финское *käsikirjoitus* ‘рукопись’ состоит из части с передним вокализмом *käsi* и части с задним вокализмом *kirjoitus*, причем суффикс имеет тот же вокализм, что и второй компонент.

§ 20. Основные речевые единицы

Как мы уже указывали, выделение морфов — вещь очень сложная. Следует отметить, что для проведения конструктивных и фонологических границ не требуется, собственно говоря, никаких сведений о морфемах, мы пользовались этим термином не при определении границ, но лишь при объяснении примеров. Поэтому можно выделить основные единицы, которые подлежат анализу, не прибегая к понятию морфемы.

Мы будем говорить, что некоторая сплошная последовательность фонем в грамматически правильной фразе составляет м и н и м а л ь н ы й с е г м е н т, если она находится между двумя ближайшими границами, выделенными по конструктивному критерию.

Если граница минимального сегмента соответствует какой-то фонологической границе, т. е. его начало и конец совпадают

с фонологическим пограничным сигналом, то такой минимальный сегмент назовем фонологически правильным.

О п р е д е л е н и е. Основной речевой единицей (сокращенно ОРЕ) является:

1) всякий фонологически правильный минимальный сегмент;

2) всякая последовательность фонем, имеющая то же распределение, т. е. встречающаяся в тех же окружениях (ср. § 23), что и некоторая ОРЕ, и неразлагающаяся на другие ОРЕ;

3) всякая последовательность фонем, которая находится между двумя ближайшими ОРЕ.

Это определение построено, как легко заметить, рекурсивно, т. е. возвращает нас каждый раз к некоторому исходному шагу, но именно так и должны, по-видимому, строиться конструктивные лингвистические определения⁵ (при построении этого определения автор опирался на некоторые соображения, высказанные устно П. С. Кузнецовым).

Легко проверить, что для русского и других славянских языков (в отличие, например, от немецкого или эстонского) ОРЕ, как правило, совпадает с тем, что обычно понимаем под «словом в речи», или «словоформой». Поскольку выделение ОРЕ в речи, как мы видим, подчинено весьма простым условиям, а ОРЕ практически для нас совпадает со словом, то мы и будем в дальнейшем оперировать понятием слова (этот термин больше способствует лингвистической интерпретации, чем ОРЕ), а понятие морфемы будем использовать лишь как производное, когда покажем, что деление на морфемы может быть получено исходя из деления на слова.

Итак, мы будем при интерпретации модели принимать в качестве исходного элемента слово (слово в речи, или словоформу). Этот выбор может быть подкреплён следующим соображением. Как мы указали в § 6, необходимо учитывать связи между моделированием и статистическим анализом языка. В частности, модель разумно строить таким образом, чтобы на ее основе могла быть построена, а затем плодотворно интерпретирована некоторая вероятностная модель. Поэтому при выборе единицы нельзя не учитывать следующего обстоятельства. Наиболее интересными в лингвистической статистике оказались именно те исследования, в которых в качестве исходной единицы бралось слово («от пробела до пробела»). Именно для слов в этом понимании был обнаружен замечательный закон Ципфа:

$$f_r = K \cdot r^{-\gamma},$$

⁵ См.: И. И. Р е в з и н. О логической форме лингвистических определений. «Применение логики в науке и технике». М., 1961.

где r — ранг слова, т. е. его номер в списке, расположенном по убывающей частотности, f_r — частотность слова с рангом r , K и γ — некоторые параметры ⁶.

Ясно, что и при статистических исследованиях непосредственным толчком для выбора слова в качестве единицы явилась его выделяемость, но значительность полученных результатов представляется все же дополнительным аргументом в пользу выбора слова в качестве исходной единицы.

Не следует, однако, думать, что при интерпретации моделей можно вообще обойтись без морфемного анализа, т. е. проведения границ внутри словоформ. Речь идет лишь о последовательности операций. Мы выбрали путь, при котором слово берется в качестве исходного понятия, а принципы морфемного анализа обсуждаются на основе слов. В некоторых работах (например, в упоминавшейся книге Гринберга) избран другой путь: сначала определяется на основе анализа фраз морфема, а затем уже определяется слово.

Строгое выделение границ внутри словоформы чрезвычайно важно и с типологической точки зрения. Дело в том, что все предложенные до сих пор типологические классификации, включая и наиболее совершенную из них, принадлежащую Сэпиру ⁷, исходят в сущности из членимости слова на части: корень, деривационные ферманты, флексии. В последнее время Гринберг предложил интересную модификацию схемы Сэпира позволяющую применять при сравнении текстов на разных языках количественные показатели, а именно следующие:

- 1) степени синтетичности: отношение $\frac{\text{число морфем}}{\text{число слов}}$;
- 2) » агглютинативности: отношение $\frac{\text{число слов с хотя бы одной границей}}{\text{число границ}}$;
- 3) » компактности (compounding): отношение $\frac{\text{число корневых морфем}}{\text{число слов}}$;
- 4) » деривативности: отношение $\frac{\text{число деривационных фермантов}}{\text{число слов}}$;

⁶ О лингвистическом значении параметров в формуле Ципфа см., например: Р. М. Фрумкина. К вопросу о так называемом «законе Ципфа». — ВЯ, 1961, № 2, стр. 118 и сл.; см. также: Д. М. Сегал. Некоторые уточнения вероятностной модели Ципфа. «Малинный перевод и прикладная лингвистика», 1961, № 5, стр. 51—55.

⁷ Эд. Сэпир. Язык. Перевод с англ. М.—Л., 1934, глава VI.

- 5) степени флективности: отношение $\frac{\text{число флексий}}{\text{число слов}}$;
- 6) » префиксации: отношение $\frac{\text{число префиксов}}{\text{число слов}}$;
- 7) » суффиксации: отношение $\frac{\text{число суффиксов}}{\text{число слов}}$

и некоторые другие.

Насколько интересны эти показатели (полученные, правда, на сравнительно коротких текстах), показывает приведенная таблица, представляющая собой некоторое сокращение таблицы Гринберга ⁸.

Показатель	Санскрит	Англосаксонский	Персидский	Английский	Якутский	Суахили	Вьетнамский
Синтетичности . . .	2,59	2,12	1,52	1,68	2,17	2,55	1,06
Агглютинативности	0,09	0,11	0,34	0,30	0,51	0,67	0
Компактности . . .	1,13	1,00	1,03	1,00	1,02	1,00	1,07
Деривативности . .	0,62	0,20	0,10	0,15	0,35	0,07	0
Флективности . . .	0,84	0,90	0,39	0,53	0,82	0,80	0
Префиксации . . .	0,16	0,06	0,01	0,04	0	1,16	0
Суффиксации . . .	1,18	1,03	0,49	0,64	1,15	0,41	0

Ясно, однако, что успешное осуществление метода Гринберга целиком зависит от наличия строгих критериев выделения единиц. В дальнейшем будут разобраны некоторые модели, позволяющие уточнить сегментацию словоформ. Пока отметим, что к показателям Гринберга вполне можно добавить отношение

$$\frac{\text{число ОРЕ}}{\text{число словоформ}}$$

Как мы видели, ОРЕ выделяются достаточно объективно. Поскольку этот показатель близок к единице для целого ряда языков, например для славянских, то, казалось бы, введение ОРЕ для типологии в том смысле, как это предлагает Гринберг, несущественно. Однако можно во всех показателях Гринберга заменить «число слов» «числом ОРЕ». Думается, что в этом случае получатся достаточно объективные показатели, интересные и с типологической точки зрения.

⁸ J. H. Greenberg. A quantitative approach to the morphological typology of languages.— IJAL, vol. XXVI, Juli 1960, № 3, стр. 178—194.

§ 21. Понятие отмеченной фразы

Мы в предыдущем изложении несколько раз пользовались понятием фразы. Термин «фраза» в лингвистике употребляется неоднозначно. Нам важно, однако, следующее. Те лингвисты, которые различали термины «фраза» и «предложение», вкладывали в термин «фраза» более широкий смысл. Особенно важно с этой точки зрения определение С. Карцевского: «Фраза — актуализованная единица коммуникации. Она не имеет собственной грамматической структуры, но обладает особой звуковой структурой, состоящей в ее интонации. Именно интонация дает фразу (*c'est précisément l'intonation qui fait la phrase*). Любое слово и объединение слов (*assemblage de mots*), любая грамматическая форма, любое междометие могут, если того требует ситуация, служить единицами коммуникации»⁹.

Посмотрим теперь, как отражается «фраза» в этом понимании в письменном тексте, который и будет предметом исследования в последующей части работы. Интонация в тексте почти не отражена (за исключением некоторых знаков препинания). Остается любое сочетание слов, вне зависимости от их грамматического оформления. Мы видим, таким образом, что «фраза» вполне может служить интерпретацией понятия «кортеж» (см. § 2) на том уровне, где единицами являются слова. Итак, фраза — это любая упорядоченная последовательность слов.

Примечание. Ясно, что отвлечение от интонационной структуры фразы есть лишь некоторая абстракция. Однако можно представить себе, что при анализе реальных языков каждой фразе приписывается еще одно абстрактное слово, состоящее из последовательности так называемых суперсегментных символов. Например, английская фраза *I'm going home* может быть записана: *I'm going home + 231 ↘*, где 2 — высокий тон, 3 — средний, 1 — сверхвысокий, а знак ↘ есть завершитель предложения (*clause terminal*). При таком подходе весь дальнейший аппарат, по-видимому, может быть использован и для случая, когда в рассмотрение вводится и интонационная структура фразы. В частности, для моделей, рассматриваемых в главе V, предусмотрена символика, позволяющая отразить и преобразование интонационных структур¹⁰.

Ясно, что принятое нами употребление термина «фраза» связано с некоторыми психологическими трудностями, которые нужно преодолеть для лучшего понимания дальнейшего. В частности, фразами будут считаться не только такие бессмысленные, хотя и грамматические построения, как

1) *кентавр выпил круглый квадрат*¹¹; 2) *идея яростно спит*,

⁹ S. Karcevskij. Sur la phonologie de la phrase.— TCLP, Prague, 1931, № 4, стр. 190.

¹⁰ См.: R. P. Stockwell. The place of intonation in generative grammar of English. «Language», vol. 36, 1960, № 3.

¹¹ См.: А. А. Реформатский и др. Введение в языкознание. М., 1960.

но и такие последовательности, как

3) *сплю идея яростно*, 4) *пришел с*.

Ясно, что все фразы (в этом понимании) распадаются на две группы: правильно построенные и неправильно построенные. Н. Хомский¹², у которого заимствованы некоторые из приведенных примеров, выдвинул следующий критерий правильности построения фразы. Фразы 1 и 2 произносятся с нормальной интонацией повествовательного предложения, в то время как 3 или 4 не может быть произнесено с такой интонацией.

Итак, Хомский, в сущности, возвращается к критерию Карцевского. Это показательно, но в известном смысле суживает интерпретации нашей модели. Неясно, почему нужно заранее отбрасывать такую вполне осмысленную интерпретацию заданных нам коротких, как отрывков текста, действительно встретившихся исследователю. Это тем более важно, что в следующих двух главах отчасти будет моделироваться ситуация, рассматриваемая в дескриптивной лингвистике. Между тем здесь термин «фраза» часто употребляется в смысле «высказывание или часть высказывания, ограниченные паузами»¹³. Под «высказыванием» (utterance) в дескриптивной лингвистике понимается любой отрезок текста, а единственной данностью для лингвиста является множество высказываний, которое Харрис именуется «универсумом речи» (см. § 3). Сюда входят лишь фразы, действительно встретившиеся исследователю, т. е. некоторое конечное множество фраз. Ниже (§ 45) мы увидим, что подобная интерпретация полезна при решении некоторых логических вопросов. Итак мы будем иметь дело с двумя интерпретациями понятия «отмеченная фраза»:

Интерпретация 1. Множество отмеченных фраз есть множество всех фраз, реально встретившихся в достаточно длинном тексте исследуемого языка (ср. интерпретацию 1 понятия «фонетическое слово» в § 9).

Интерпретация 2. Множество отмеченных фраз есть множество грамматически правильных фраз исследуемого языка или, что, по-видимому, то же самое, множество потенциально возможных фраз данного языка (ср. интерпретацию 2 понятия «фонетическое слово» в § 9).

Интерпретация 2 будет в дальнейшем основной. Поскольку, однако, понятие грамматической правильности является интуитивным, то мы будем считать, что критерий отмеченности лежит вне формальной системы. Можно представить себе дело так, что имеется «оракул», к которому можно обратиться для проверки того, является ли некоторая фраза отмеченной.

¹² См.: Н. Хомский. Указ. соч., стр. 418—420; см. также рецензию Е. В. Падучевой.— ВЯ, 1959, № 1.

¹³ См.: В. Вloch. Studies in colloquial Japanese, IV. Phonemics. «Language», 1950, № 26, стр. 90.

§ 22. Понятие В-структуры фразы

Мы будем рассматривать в дальнейшем всевозможные разбиения множества исходных элементов, т. е. слов (словоформ) на непересекающиеся классы. С подобными классами мы встречаемся в любой грамматической теории. Можно представить себе, что все слова разбиты на классы существительных, прилагательных, глаголов и т. п., или же все слова разбиты на классы по их окончанию, или, например, по типу основы (как это принято для древних индоевропейских языков) и т. п.

Интересная классификация возникающих в лингвистике классов и подклассов слов предложена А. А. Холодовичем¹⁴. Нас, однако, сейчас не интересует критерий разбиения на классы. Важно лишь, что по какому-то критерию все слова разбиты на непересекающиеся классы.

Каждой фразе исследуемого языка может быть теперь поставлена в соответствие цепочка классов принятого нами разбиения. Например, если нам задано разбиение на класс существительных (S), класс прилагательных (A), класс предлогов (Pr), класс глаголов (V) и класс остальных слов (Z), то, например, отмеченной фразе:

Маленький мальчик не видит меня и девочки

соответствует цепочка классов:

A S Z V Z Z S.

Этой же цепочке классов может соответствовать и неотмеченная фраза, например:

Красную слон хотя бегать меня но собак.

Можно, наконец, придумать цепочку классов, которой не соответствует ни одна отмеченная фраза, например:

Pr Pr Pr Z Pr.

Рассмотрим теперь ситуацию, возникающую при некотором абстрактном разбиении В. Пусть дана фраза: $A = x_1 x_2 \dots x_n$.

Сопоставим с каждым словом x_i класс В (x_i) при данном разбиении В. Иначе говоря, устроим отображение множества слов на множество классов. Цепочка классов $V(x_1) V(x_2) \dots V(x_n)$, соответствующая данной фразе А при данном разбиении В, будет называться В-структурой фразы А и обозначаться $V(A)$. В-структура называется отмеченной, если хотя бы одна фраза, соответствующая данной структуре, является отмеченной¹⁵.

¹⁴ А. А. Холодович. Опыт теории подклассов слов.— ВЯ, 1960, № 1.

¹⁵ Эти понятия, как и большинство понятий, используемых в дальнейшем, сформулированы А. А. Ляпуновым и О. С. Кулагиной, см.:

В-структура называется с о в е р ш е н н о й ¹⁶, если любая фраза, соответствующая данной структуре, является отмеченной.

Примечание. Использование термина «структура» в смысле цепочки некоторых множеств на первый взгляд противоречит принятому лингвистическому словоупотреблению. Но слово «структура» в наше время употребляется столь часто, что перестает нести какую-нибудь информацию. Поэтому вполне допустимо придать ему то значение, которое оно имело в период оформления современной лингвистики и современной логики. Наше употребление термина «структура», взятое у О. С. Кулагиной, вполне соответствует употреблению этого термина в следующем высказывании: «Проницательная логистика нашего времени знает и употребляет в своей области с замечательным успехом понятие „структуры“ (den Strukturbegriff); можно без особых усилий в таком простом латинском предложении, как *Caius amavit Camillam*, осуществить абстракцию, которая позволяет увидеть то, что логистика называет структурным; я пишу -us, -avit, -am и представляю себе пустые места заполнимыми другими словами соответствующего класса»¹⁷

§ 23. Понятие эквивалентности

При всем различии, наблюдаемом между отдельными школами в современной лингвистике, имеется нечто, объединяющее все эти направления, а именно метод дистрибутивного анализа ¹⁸, состоящий в выявлении классов элементов, взаимозаменяемых в том или ином смысле. Процедура проверки на взаимозаменяемость, называемая в копенгагенской школе «тестом коммутации», а в дескриптивной лингвистике «субституцией» (подстановкой), является, по-видимому, одним из основных инструментов исследования в современной лингвистике. Особое значение этот прием приобретает в дескриптивной лингвистике, в особенности для тех ее представителей, которые полностью отказываются от изучения значений, а всю лингвистическую структуру сводят к распределению фонем и морфем в данном языке ¹⁹. Целесообразно, однако, не связывать деятельность представителей этого направления с их взглядами на сущность, границы и общие методы лингвистического исследования, а закрепить термин «дескриптивная лингвистика» за тем разделом науки о языке, который занимается анализом

О. С. Кулагина. Об одном способе определения грамматических понятий на базе теории множеств. «Проблемы кибернетики», вып. 1. М., 1958, стр. 203—214.

¹⁶ Идея «совершенности» принадлежит Дж. Гринбергу; он, правда, пользуется термином «совершенная конструкция» (вернее «подконструкция»), но это не меняет сути дела.

¹⁷ К. Bühler. *Das Strukturmodell der Sprache*.— TCLP, Prag, 1936, 6, стр. 10.

¹⁸ См.: Э. Хауген. Направления в современном языкознании. «Новое в лингвистике», вып. 1. М., 1960, стр. 249—251.

¹⁹ См., например: С. F. Voegelin. *Linguistics without meaning and culture without words*. «Word», t. V, 1949, № 1, стр. 40 и сл.

отношений между некоторыми лингвистическими объектами²⁰ на основе анализа распределения, т. е. пользуясь методами субституции.

Интересно, что идущая от античных грамматик и осмеянная в теоретических курсах синтаксиса традиция постановки вопросов типа «*Кто* сидит в комнате?» для выявления подлежащего на самом деле при правильном понимании сути дела, давно уже потерянным в грамматике, очень близка к идеям дистрибутивного анализа. Как отмечает А. М. Пешковский, вопросы суть «один из видов подстановки в словосочетания одних слов на месте других»²¹. Выделяется класс слов, эквивалентных с той точки зрения, что все они могут стоять вместо многозначия во фразе «...сидит в комнате».

Аналогично соответствующая традиция в логике идет от Аристотелева учения о категориях²². И здесь одна из линий развития привела к «семантической категории» (*Bedeutungskategorie*) Гуссерля²³, а именно к понятию класса слов или групп слов такого, что замена одних элементов класса другими не приводит к обесмысливанию всего контекста. Дальнейшее уточнение это понятие получило в работе Айдукевича о синтаксической связности²⁴.

Анализ проблем синонимии в реальных и искусственных языках привел Карнапа²⁵ к понятию «экстенционала» некоторого предиката, а именно совокупности предметов, к которым применим данный предикат. Два предиката имеют одинаковый экстенционал, если они применимы к одинаковым предметам. С лингвистической точки зрения речь идет об анализе всех возможных фраз некоторого фиксированного языка, в которых встречаются предикаты *P* и *Q*.

Эти замечания показывают, что значение идеи эквивалентности выходит далеко за рамки строго дескриптивного анализа. Мы начнем, однако, с тех понятий, которые используются в дескриптивной лингвистике. Будем считать, что два слова *x*

²⁰ См.: Вяч. Вс. Иванов. О построении информационного языка для текстов по дескриптивной лингвистике. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста», вып. 7, М., 1961.

²¹ А. М. Пешковский. Вопрос о «вопросах». — «Избранные труды». М., 1959, стр. 35.

²² См.: G. Ryle. *Categories*. «Logic and Language», 2. Series. Oxford, 1959, стр. 66—67.

²³ E. d. Husserl. *Logische Untersuchungen*. Bd 2, T. 1. Zweite, umgearbeitete Auflage. Halle a. d. S., 1913, стр. 294—295, 305—312, 316—321, 326—342.

²⁴ К. Айдукевич. Die syntaktische Konnexität. «*Studia philosophica*», vol. 1. Lwów, 1935.

²⁵ Р. Карнап. Значение и необходимость. Перевод с англ. М., 1959, стр. 57 и сл.; там же см. ссылки на другие работы, в которых ставится проблема взаимозаменяемости слов.

и *y* относятся к одному классу относительно данного контекста, или, короче, к одному «контекстуальному классу», если встретились две отмеченные фразы, в которых *x* и *y* находились в одинаковом окружении. Например, к одному контекстуальному классу относятся в русском языке слова *слона*, *дом*, *селёдку*, поскольку все они встречаются во фразе «я вижу...». Ясно, что каждое слово входит в несколько контекстуальных классов.

Понятие контекстуального класса интересно тем, что, по-видимому, позволяет переходить от интерпретации 1 множества отмеченных фраз к интерпретации 2 (т. е. интерпретации множества отмеченных фраз как правильно построенных, или потенциальных фраз).

Пусть реально встретившиеся, т. е. отмеченные по интерпретации 1, фразы нам заданы. Тогда можно предложить следующее определение правильно построенной, или потенциальной, фразы. Будем считать, что фраза правильно построена, если она отмечена по интерпретации 1 или получена из правильно построенной фразы путем замены одного слова *x* словом, принадлежащим к одному из контекстуальных классов, в которые входит *x*.

Покажем, например, что фраза *Идея яростно спит* построена правильно. В самом деле, фраза *Девочка спокойно спит* отмечена по интерпретации 1 и потому правильно построена (первое условие в нашем определении).

Слова *равнодушно* и *спокойно* относятся к одному контекстуальному классу (по контексту «он...сказал»), слова *равнодушно* и *яростно* относятся к одному контекстуальному классу (по контексту «...взглянуть», ср. примеры, приводимые в словаре Ожегова на слово *взглянуть*). Слова *девочка* и *идея* относятся к одному классу (по контексту «счастливая...», ср. *счастливая идея* в словаре Ушакова).

Получаем последовательность фраз:

Девочка спокойно спит,
Девочка равнодушно спит,
Девочка яростно спит,
Идея яростно спит,

причем все фразы, в том числе и последняя, построены правильно.

Это построение очень напоминает переход от интерпретации 1 фонетического слова к интерпретации 2 (см. § 9), и в этом его основной интерес.

Однако контекстуальные классы неудобны для операций внутри модели (для которой множество отмеченных фраз задается извне), поскольку, во-первых, эти классы пересекаются для разных контекстов и, во-вторых, в эти классы могут вхо-

дить и случайные слова, например, *хорошо* входит в тот же контекстуальный класс, что и *слона*, *стол*, *селедку* (по контексту «я вижу...»). Поэтому в моделях, которые мы рассмотрим, понятие взаимозаменяемости формулируется так, чтобы оно разбивало все слова на непересекающиеся классы.

Этому требованию отвечает следующее определение эквивалентности. Мы будем говорить, что элемент x_1 *Э*-эквивалентен элементу x_2 (записывается $x_1 \tilde{E} x_2$), если 1) для любой отмеченной фразы вида $A_1 x_1 A_2$ фраза $A_1 x_2 A_2$ также является отмеченной и 2) для любой отмеченной фразы $B_1 x_1 B_2$ фраза $B_1 x_2 B_2$ также является отмеченной (A_1, A_2, B_1 и B_2 — произвольные фразы ²⁶).

В этой форме определение эквивалентности сформулировано В. А. Успенским ²⁷. У О. С. Кулагиной формулировка имеет несколько другой вид: $x \tilde{E} y$, если для любых фраз A и B фразы вида $A x B$ и $A y B$ одновременно отмечены или одновременно не отмечены. Ясно, что обе формулировки равноценны, и мы будем в дальнейшем пользоваться как той, так и другой.

E-эквивалентность обладает следующими свойствами:

1) рефлексивность: $x \tilde{E} x$ (для любой отмеченной фразы $A_1 x A_2$ фраза $A_1 x A_2$ отмечена);

2) симметричность: если $x \tilde{E} y$, то $y \tilde{E} x$ (симметричность вытекает из симметричности первого и второго условий);

3) транзитивность: если $x \tilde{E} y$ и $y \tilde{E} z$, то $x \tilde{E} z$ (если $A_1 x A_2$ отмечена, то отмечена $A_1 y A_2$, если $A_1 y A_2$ отмечена, то отмечена $A_1 z A_2$, т. е. если отмечена $A_1 x A_2$, то отмечена $A_1 z A_2$; точно так же показывается, что из отмеченности $B_1 z B_2$ следует отмеченность $B_1 x B_2$).

Все множество слов разбивается на ряд непересекающихся подмножеств, которые мы будем называть семействами (S_1, S_2, \dots, S_n).

Если множество отмеченных фраз интерпретировать как множество правильно построенных фраз русского языка, то мы получим следующие примеры семейств:

S_1 — *столу, окну, человеку, клопу, дуралею...*

S_2 — *хохотал, прыгал, бегал, визжал...*

S_3 — *хорошо, плохо, сомнительно...*

S_4 — *когда, где, как...*

²⁶ Напомним, что фразой называется любая последовательность слов и, например, любая часть обычного предложения также есть фраза. Наконец, для общности удобно считать, что фраза может быть пустой, т. е. не содержать ни одного слова.

²⁷ В. А. Успенский. К определению части речи в теоретико-множественной системе языка. — «Бюллетень объединения по проблемам машинного перевода», № 5. М., 1957.

Каждой фразе можно теперь сопоставить последовательность семейств, так чтобы каждому элементу соответствовало определенное семейство. Такую последовательность мы будем называть *Структурой фразы* *A* и обозначать *S(A)*.

Фраза	<i>Профессор</i>	<i>вошел</i>	<i>в</i>	<i>аудиторию</i>
<i>S-структура</i>	<i>профессор</i>	<i>вошел</i>	<i>в</i>	<i>аудиторию</i>
	<i>слон</i>	<i>прыгнул</i>	<i>на</i>	<i>комнату</i>
	<i>человек</i>	<i>двинулся</i>	<i>и т. д.</i>	<i>кухню</i>
	<i>сторож</i>	<i>приехал</i>		<i>звезду</i>
	<i>предок</i>	<i>и т. д.</i>		<i>комету</i>
	<i>и т. д.</i>		<i>ракету</i>	<i>и т. д.</i>

Очень напоминает понятие *S-структуры* знаменитый пример, построенный Л. В. Щербой: *Глокая куздра штенно будганула бокра и кудлявит бокренка*. Это не есть фраза русского языка, ибо ни одно из слов не является словом русского языка. Но она может в известном смысле рассматриваться как *S-структура*, а именно: *глокая* — как название семейства всех эквивалентных между собой прилагательных на *-ая*: *красивая, большая, глубокая, пестрая* и т. п.; *куздра* — как название семейства всех эквивалентных между собой существительных ж. р. ед. ч.: *слониха, стрекоза, корова, мамаша* и т. п.; *штенно* — как название семейства наречий образа действия: *хорошо, плохо, сильно, слабо* и т. п.; *будганула* — как название семейства глаголов типа *кольнула, ущипнула, подковырнула* и т. п. *S-структура «Глокая куздра...»* является, разумеется, отмеченной, ибо имеется, например, фраза русского языка: *Грузная слониха ловко подтолкнула слона и ласкает слоненка*, которая является отмеченной.

Приведенные примеры могут привести к созданию ложного впечатления, что семейство соответствует классу слов с фиксированной формой. В общем случае это далеко не так. Например, существительные индонезийского языка, ничем не различающиеся по оформлению, распадаются на такие семейства:

S_1 — <i>mahasiswa</i>	‘студент’	<i>dewi</i>	‘богиня’
<i>murid</i>	‘ученик’		
S_2 — <i>harimau</i>	‘тигр’	<i>manuk</i>	‘курица’
<i>kuda</i>	‘лошадь’		
S_3 — <i>rumah</i>	‘дом’	<i>pisang</i>	‘банан’
<i>buku</i>	‘книга’		
S_4 — <i>surat</i>	‘письмо’	<i>bedil</i>	‘ружье’ и т. д.

Группировка в семейства в этом языке связана с употреблением так называемых счетных слов. Слова из S_1 это те, которые могут стоять после счетного слова *orang* 'человек', слова из S_2 стоят после счетного слова *ekor* 'хвост', слова из S_3 после счетного слова *buah* 'плод' и слова из S_4 после слова *putjuk* 'отросток'. Речь идет не о семантических, а именно грамматических группировках, поскольку соответствующие счетные слова обязательно употребляются после каждого числительного, например:

saja membeli manuk tiga ekor
'я' 'купил' 'курица' 'три' 'хвост'

В других языках, например в суахили, подобные — на первый взгляд семантические — группировки в семейства оформлены и морфологически. Так, для суахили семейства соответствуют грамматическому классу.

Распределение по семействам связано не с оформлением слова, хотя в ряде языков, например в славянских языках, оно и сопровождается определенными морфологическими показателями. Оно связано с синтаксическими свойствами слова и зависит от того, в каких (грамматических) контекстах употребляется данное слово. В этом ценность понятия «семейство», которое применимо к анализу произвольных языков вне зависимости от той морфологической техники, которая характерна для языков определенного типа.

Заметим теперь, что возможен и другой подход к понятию семейства, приводящий к тем же самым результатам. Вместо эквивалентности слов можно рассматривать эквивалентность о к р у ж е н и й (environment) и выделять класс эквивалентных окружений для некоторого слова. Эта идея была использована А. Н. Колмогоровым для построения формального аналога «падежа».

Определение падежа по Колмогорову известно в изложении В. А. Успенского²⁸. В этом изложении используются исходные понятия «предмета» и «состояния», которые не могут быть определены в терминах, введенных в § 20 и § 21.

Мы попытаемся изложить идею определения, не выходя за пределы того, что может быть сформулировано в наших исходных терминах.

Назовем «фразой с многоточием» (Р. Л. Добрушин) отмеченную фразу, в которой одно слово заменено многоточием.

Две фразы с многоточием назовем эквивалентными относительно некоторого слова *x*, если в обеих вместо многоточия

²⁸ В. А. Успенский. К определению падежа по Колмогорову. «Бюллетень объединения по проблемам машинного перевода», № 5. М., 1957, стр. 22—26.

можно подставить x так, что получившаяся фраза будет отмечена. Например, в русском языке фразы с многоточием «...*кипит*» и «*кошка пьет...*» эквивалентны относительно слова *молоко*. Эти же фразы с многоточием не эквивалентны относительно слова *вода*. В связи с этим введем новое определение. Две фразы с многоточием абсолютно эквивалентны, если они эквивалентны относительно любого слова, которое при подстановке его хотя бы в одну из двух фраз с многоточием делает фразу отмеченной. Например, фразы с многоточием «*кошка пьет...*», «*кошка любит...*», «*я вижу...*» абсолютно эквивалентны. Абсолютная эквивалентность двух фраз с многоточиями обладает свойствами симметрии, рефлексивности и транзитивности. Все множество фраз с многоточиями разбивается на непересекающиеся классы. Эти классы А. Н. Колмогоров предложил называть падежами.

Легко видеть, что эти классы соответствуют семействам. В § 34 мы увидим, что соответствие между падежами и семействами достаточно сложно. Но сама идея определения традиционных лингвистических категорий в терминах эквивалентности оказалась, как мы увидим в главе IV, весьма плодотворной.

Обобщим теперь понятие эквивалентности на произвольные В-структуры.

Будем считать, что элемент V_1 (при данном разбиении V) В-эквивалентен элементу V_j и запишем это $V_1 \sim V_j$, если для любой отмеченной структуры вида $V(A_1)V(A_2)$ структура $V(A_1)V_jV(A_2)$ отмечена и для любой отмеченной структуры $V(D_1)V_j(D_2)$ структура $V(D_1)V_1V(D_2)$ отмечена.

Если мы сравним это определение с определением Е-эквивалентности, то мы видим, что там мы имеем дело просто с частным случаем В-разбиения, а именно таким, когда каждый класс состоит ровно из одного слова; такое разбиение мы будем называть е д и н и ч н ы м.

§ 24. Понятие окрестности

В целом ряде моделей будет использовано понятие окрестности. Будем считать, что кроме исходных единиц и некоторых отмеченных кортежей нам задано некоторое разбиение на непересекающиеся классы, которые мы будем называть окрестностями, т. е. для каждого элемента x известно, в какую окрестность $\Gamma(x)$ входит данный элемент.

Прежде всего возникает вопрос, почему в грамматических моделях появляется необходимость в некотором новом исходном понятии, которое не имеет аналога в фонологических моделях. Связано это с одним кардинальным отличием грамматики от фонологии, которое состоит, как известно, в следующем: последняя имеет дело лишь с п л а н о м в ы р а ж е н и я в

языке, т. е. с некоторой совокупностью формальных сигналов, в то время как в грамматике мы не можем целиком отвлечься от плана содержания, т. е. должны учитывать не только означающие, но и означаемые.

С другой стороны, ясно, что в модели план содержания должен быть отражен сугубо формальным способом как нечто заданное. Это и достигается заданием окрестностей.

Принадлежность к одной окрестности двух слов (ОРЕ) отражает то, что они несут одинаковую семантическую информацию. Последнее, однако, не очень точно, поэтому в дальнейшем мы поясним на примерах, какие интерпретации целесообразно дать понятию окрестности.

Основная интерпретация понятия окрестности состоит в следующем: это система форм словоизменения одного и того же слова, например, *стол, стола, столу, столами, столах* и т. п., принадлежат одной окрестности, а *стол, столик, столовый, столоваться* принадлежат попарно к разным окрестностям. Ясно, конечно, что этот вопрос в лингвистике не решен однозначно²⁹. Хотя построения внутри модели никоим образом не зависят от выбора интерпретации понятия окрестности, мы будем иметь в виду структурный принцип регулярности систем словоформ, а внутри тех случаев, на которые распространяется этот принцип, выделим две интерпретации.

Интерпретация 1. Внутри одной окрестности попадают те и только те формы, объединенные общей лексической морфемой, которые входят в одну парадигму склонения или в одну парадигму спряжения.

Эту интерпретацию мы будем, однако, иметь в виду лишь в тех параграфах, где строится аналог «части речи». Однако понятие части речи имеет, по-видимому, подлинную ценность лишь там, где слова выступают в своей «первичной» функции³⁰, т. е. там, где существительные обозначают предметы, прилагательные обозначают постоянные признаки, глаголы — переменные признаки и т. п. Поэтому когда мы в дальнейшем будем строить аналог «части речи», то мы будем иметь в виду лишь слова (в первую очередь существительные) в их первичной функции.

Необходимо признать, что не очень точен наш критерий вхождения в одну парадигму склонения или спряжения. «Сами понятия „склонения“ и „спряжения“ никогда не были точно

²⁹ См. анализ различных точек зрения в кн.: В. В. Виноградов. Современный русский язык. Грамматическое учение о слове, вып. 1: М., 1938.

³⁰ Ср. понятие первичной синтаксической функции, введенное Ю. Куриловичем (J. Kurylowicz. *Dérivation lexicale et dérivation syntaxique. Contribution à la théorie des parties du discours.* «BSL de Paris», t. 37. 1936, стр. 79—92.)

определены, а взяты прямо из традиции»³¹. Исходя из этого, мы, за исключением § 27 и 28, будем иметь в виду под «окрестностью» другое, а именно совокупность слов таких, которые связаны общей лексической морфемой и если для одного из них, имеющего вид

$$A + B$$

основа суффикс + окончание

нашлось слово вида

$$A' + B,$$

то для любого другого, имеющего вид

$$A + B',$$

найдется слово вида

$$A' + B'.$$

Иначе говоря, мы будем иметь в виду слова, объединяемые между собой по принципу тождества основы так называемым методом квадратов (ср. § 31). Таким образом, мы получаем следующую интерпретацию понятия «окрестность».

Интерпретация 2. Внутрь одной окрестности попадают формы, имеющие общую основу и образованные наиболее продуктивными флексиями и суффиксами, почти не знающими ограничений в своем употреблении.

Таким образом, в окрестность некоторой глагольной формы попадут причастия, инфинитив, деепричастия, *nomina agentis* и *nomina actionis*.

В некоторых славянских языках, например в чешском и словацком, в окрестность имени, обозначающего одушевленное существо, следует при второй интерпретации отнести притяжательные прилагательные. Очень интересно мнение Н. С. Трубецкого относительно подобных прилагательных старославянского языка:

«От каждого существительного, обозначающего одушевленное существо, образуется притяжательное прилагательное, которое принадлежит к парадигме склонения этого существительного, так же как и причастия принадлежат к парадигме спряжения глагола»³².

Эта интерпретация и будет иметься в виду в большей части работы. Учитывая, однако, что каждая из наших интерпретаций связана со значительными трудностями, мы примем еще одну интерпретацию.

³¹ А. М. Пешковский. Понятие отдельного слова. «Сборник статей». Л., 1925, стр. 131.

³² Н. С. Трубецкой. О притяжательных прилагательных (*possessiva*) старосербо-славянского языка. «Сборник у част Белића». Београд, 1937, стр. 16.

Интерпретация 3. Внутри одной окрестности попадают все слова с общей лексической морфемой (лексемой) без каких-либо ограничений на продуктивность окончаний и суффиксов.

В § 30 и 49 мы покажем, как внутри самой модели могут быть сформулированы критерии для перехода от интерпретации 3 к интерпретации 2.

Необходимо сделать еще одно замечание относительно понятия окрестности. В начале данного параграфа было сказано, что разбиение на окрестности задается как разбиение на непересекающиеся классы. Между тем, если иметь в виду интерпретации 1 и 2, то системы форм двух разных слов могут пересекаться:

I	II
<i>печь</i>	<i>печь</i>
<i>пеку</i>	<i>печи</i>
<i>печешь</i>	<i>печью</i>
<i>пек</i>	. . .
<i>пекли</i>	<i>печей</i>
.

Мы уже указали, что можно исходить из интерпретации 3, при которой мы будем иметь всего одну окрестность, связанную общей морфемой *пек/печь*.

Но интерпретации 1 и 2 весьма важны с точки зрения лингвистического истолкования формальных построений ряда параграфов. И можно считать, что грамматические омонимы типа *печь* заранее заданы нам как два слова (*печь₁* и *печь₂*) и подобными индексами снабжены все омонимы. Далее мы покажем, что существуют и формальные модели для различения омонимов.

§ 25. Грамматические категории

Последнее понятие, которое в дальнейшем будет считаться заданным извне, т. е. неопределяемым в терминах модели, есть понятие грамматической категории (иногда мы будем просто говорить «категория»). Иначе говоря, мы будем считать, что нам заданы некоторые абстрактные понятия, называемые грамматическими категориями³³ и для каждого слова указывается,

³³ Для интерпретации того, что следует считать грамматическим, а что неграмматическим, по-видимому, весьма плодотворна идея И. А. Мельчука, который предложил называть «грамматическим в данном языке» значение, которое в данном языке выражается обязательно, т. е. слова соответствующих классов немислимы без показателей этого значения. При этом выдвигается еще дополнительное требование, чтобы существовал хотя бы один язык, где это значение не выражается (И. А. Мельчук. К вопросу о грамматическом в языке-посреднике. «Тезисы совещания по математической лингвистике». Л., 1959, стр. 60). В дальнейшем рассматриваются только те категории, которые И. А. Мельчук назвал

с какими категориями оно ставится в соответствие. При этом одно слово ставится в соответствие, вообще говоря, нескольким категориям. Отображение множества слов на множество категорий индуцирует некоторое разбиение слов на классы, пересекающиеся между собой. Эти классы слов удобно также называть категориями.

Подобное понимание категории близко к тому, которое было предложено А. М. Пешковским. В самом деле. Для Пешковского грамматическая (или, как он говорит, формальная) категория есть множество слов, характеризующихся общим грамматическим значением и общим формальным признаком (хотя бы у части слов)³⁴. Например, следуя Пешковскому, можно выделить следующие формальные категории слов:

Им. п.	Дат. п.	Вин. п.	Ед. ч.	Мн. ч.
<i>корова</i>	<i>корове</i>	<i>корову</i>	<i>корова</i>	<i>коровы</i>
<i>коровы</i>	<i>коровам</i>	<i>коров</i>	<i>корове</i>	<i>коров</i>
<i>лампа</i>	<i>лампе</i>	<i>лампу</i>	<i>корову</i>	<i>лампы</i>
<i>лампы</i>	<i>лампам</i>	<i>лампы</i>	<i>лампа</i>	и т. д.
<i>стол</i>	и т. д.	и т. д.	<i>лампе</i>	
<i>столы</i>			<i>лампу</i>	
<i>человек</i>			<i>стол</i>	
<i>люди</i>			<i>столу</i>	
<i>окно</i>			<i>человек</i>	
<i>окна</i>			и т. д.	
<i>тело</i>				
<i>тела</i>				
<i>пальто</i>				
и т. д.				
Ж. р.	М. р.	Ср. р.	Одуш.	Неодуш.
<i>корова</i>	<i>стол</i>	<i>окно</i>	<i>корова</i>	<i>стол</i>
<i>коровы</i>	<i>столу</i>	<i>тело</i>	<i>лань</i>	<i>окно</i>
<i>коров</i>	<i>человек</i>	<i>окну</i>	<i>корову</i>	<i>столу</i>
<i>коровам</i>	<i>человеку</i>	<i>телу</i>	<i>лани</i>	<i>окну</i>
<i>лампа</i>	и т. д.	<i>пальто</i>	<i>человек</i>	<i>пальто</i>
<i>лампы</i>		и т. д.	<i>люди</i>	и т. д.
и т. д.			и т. д.	

«морфологическими», поэтому в нашей работе термин «морфологическая категория» будет употребляться в более узком значении.

³⁴ А. М. Пешковский. Русский синтаксис в научном освещении, 6-е изд. М., 1938, стр. 53—61.

Поскольку мы видим, что слово соответствует одновременно целому ряду категорий, то полезно ввести понятие, соответствующее совокупности или вернее конъюнкции нескольких категорий.

Мы будем говорить, что два слова x и y относятся к одной элементарной грамматической категории³⁵, если категории, к которым одновременно относятся x и y , совпадают, иначе говоря, если x относится ко всем категориям, к которым относится y , и y относится ко всем категориям, к которым относится x .

В нашем примере *окну* и *телу* относятся к одной элементарной категории (конъюнкция категорий: дательный падеж и единственное число и средний род и неодушевленность³⁶), а *человек* и *стол* — к разным, потому что *стол* входит в категорию «неодушевленность», а *человек* — в категорию «одушевленность»). О случаях дизъюнктивной связи элементарных категорий (например, *стол* соответствует дизъюнкции двух конъюнкций: мужской род и единственное число и неодушевленность и именительный падеж или мужской род и единственное число неодушевленность и винительный падеж) речь будет идти в § 32.

В дальнейшем нам понадобится классификация категорий по отношению этих категорий к окрестностям. Здесь полезно выделить два типа категорий.

1. Некоторые категории устроены так, что если некоторое слово x из окрестности $\Gamma(x)$ входит в эту категорию, то любое слово из $\Gamma(x)$ входит в эту категорию. Такие категории мы назовем **п а р а д и г м а т и ч е с к и м и**. Примеры парадигматических категорий существительного: мужской род, женский род, средний род, одушевленность.

2. Всякую категорию, которая не обладает указанным свойством, мы назовем **н е п а р а д и г м а т и ч е с к о й**. Примеры непарадигматических категорий существительного: множественное число, дательный падеж и т. д. (Различие парадигматических и непарадигматических категорий во многом совпадает с различием между несинтаксическими и синтаксическими категориями у Пешковского, в следующем параграфе мы увидим, что существует и другая возможность приблизиться к этим категориям Пешковского.)

³⁵ Понятие элементарной грамматической категории соответствует понятию «элементарной информации», введенному в работе: С. Я. Ф и т а л о в. О построении формальной морфологии в связи с машинным переводом. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста», вып. 2, М., 1961, стр. 2. Термин «элементарная грамматическая категория» был в несколько другом смысле употреблен Р. Л. Добрушиным (см. об этом § 32).

³⁶ По поводу категорий «одушевленность» и «неодушевленность», см. § 29.

В начале параграфа мы отметили, что грамматическая категория есть исходное понятие, задаваемое извне. Иначе говоря, оно не зависит от других исходных понятий. Для разумной лингвистической интерпретации наших моделей целесообразно, однако, чтобы задание категорий было определенным образом связано с заданием окрестностей.

Пока мы можем выдвинуть лишь требование, касающееся непарадигматических категорий, о задании числа парадигматических категорий мы будем говорить в § 29. Требование же относительно числа непарадигматических категорий состоит в следующем.

Если два слова x и y , входящие в одну окрестность, принадлежат разным семействам, то должна существовать пара непарадигматических категорий K_1 и K_2 , таких, что x входит в K_1 и не входит в K_2 , а y входит в K_2 и не входит в K_1 .

Теперь мы закончили конструкцию исходных понятий, в терминах которых строятся все дальнейшие модели.

§ 26. О так называемом изоморфизме между фонологией и грамматикой

Часто говорят об изоморфизме³⁷ между фонологической и грамматической системой. Ниже будет показано, как можно в наших терминах отразить эту проблему.

Мы условились, что нам задано некоторое множество грамматических категорий, таких, как женский род, множественное число, одушевленность и т. п., причем каждое слово ставится в соответствие с некоторым подмножеством всего множества категорий, а именно элементарной грамматической категорией.

Мы будем называть две категории *совместимыми* в данном языке, если существует хотя бы одно слово, поставленное в соответствие с элементарной категорией, которая одновременно входит в обе категории.

В противном случае мы будем говорить о *несовместимых* категориях. Примеры попарно совместимых категорий: женский род, множественное число, неодушевленность.

Мы будем называть две категории K_1 и K_2 *однородными*, если они несовместимы и существует хотя бы одно слово данного языка, такое, что замена K_1 на K_2 (или K_2 на K_1) в соответствующем ему подмножестве категорий приводит к совокупности категорий, соответствующей некоторому слову данного языка.

Примеры однородных категорий в русском языке: мужской род, женский род, средний род. Неоднородными несовместимы-

³⁷ J. Kurýlowicz. La notion d' isomorphisme, TCLC, 1949, V, стр. 56 и сл.

ми категориями являются прошедшее время и дательный падеж.

В этом параграфе мы будем оперировать не словами, а соответствующими им совокупностями, или *n*-ками категорий.

Каждой отмеченной фразе соответствует, таким образом, некоторая последовательность, состоящая из *n*-ок категорий.

Мы будем называть категорию K_1 с в я з а н н о й по отношению к категории K_2 в данной позиции, если появление K_1 обусловлено наличием K_2 в одной из предшествующих или последующих *n*-ок в последовательности, соответствующей данной фразе.

Если категория K_1 связана по отношению к K_2 в каждой позиции, то она называется *н е м о р ф о л о г и ч е с к о й*. Остальные категории называются *м о р ф о л о г и ч е с к и м и* (ср. сноску 33 на стр. 72). Примером неморфологической категории является женский род или дательный падеж прилагательных.

Если некоторая морфологическая категория связана в данной позиции, то мы будем называть данную позицию *п о з и ц и е й н е й т р а л и з а ц и и*³⁸ противопоставления данной категории и всех однородных с ней. Например, позиция после *чтобы* во фразах русского языка типа «Я хочу, *чтобы* он *пришел*» есть позиция нейтрализации категории «прошедшее время» по отношению к однородной к ней категории «непрошедшее время».

П р и м е ч а н и е. Наше определение нейтрализации более узко и, по-видимому, более формально, чем предложенное чешским лингвистом Б. Трнака. Так, приводимый им английский пример *My brothers are merchants* 'Мои братья — коммерсанты' есть, с нашей точки зрения, нейтрализация в позиции предиката, поскольку здесь категория множественного числа для слова *merchants* связана по отношению к категории множественного числа в слове *brothers*. Трнака также считает это случаем нейтрализации, но по другим причинам: «форма множественного числа *merchants* не обозначает два или более индивидуумов, как в случае управляющего существительного *brothers*, но просто понятие коммерсанта (*a merchant*) без всякой числовой квалификации»³⁹. В связи с этим некоторые примеры нейтрализации, приводимые Трнакой (например, снятие противопоставления между единственным и множественным числом во всех позициях за исключением субъекта предложения в некоторых дравидских языках), не подходят под наше определение нейтрализации.

³⁸ Здесь использованы некоторые идеи, высказанные Вяч. Вс. Ивановым и П. С. Кузнецовым на семинаре по математической лингвистике в МГУ в 1957/58 учебном году.

³⁹ В. Т р н а к а. On some problems of neutralization. «Omagiu lui Jorgu Jordan cu prilejul implinirii a 70 de ani». Bucarești, 1958, стр. 865—866.

Назовем теперь «грамматемой» любую совокупность, которая состоит из n морфологических категорий, поставленных в соответствие с некоторым словом.

Назовем «архиграмматемой» совокупность из $n-1$ категорий, совпадающих для двух грамматем.

Читатель уже заметил, что между терминами этого параграфа и фонологическими терминами главы II можно установить взаимоднозначное соответствие, сохраняющее все отношения (например, термин «грамматема» так относится к термину «морфологическая категория», как термин «фонема» к термину «релевантный признак»).

Пусть теперь все грамматемы некоторой подсистемы упорядочены способом, которым упорядочены фонемы (процедура из § 12 может быть полностью переписана с заменой слова «фонема» словом «грамматема»). Мы получим некоторую упорядоченную систему грамматем, обладающую теми же абстрактными свойствами, что и система § 12.

В частности, определим понятие «категориальной однородности», аналогичное понятию фонологической однородности; определим понятие пустой клетки так же, как в § 13 (с заменой слов «признак» и «фонема» словами «категория» и «грамматема»). Тогда, очевидно, будут иметь место утверждения, аналогичные теоремам 2.1—2.5. Например:

Теорема 3.1. Если грамматическая подсистема не имеет пустых клеток, то она категориально однородна.

Таким образом, мы видим, что изоморфизм между терминологическими системами действительно построить очень легко. Что же касается реальных объектов, то здесь положение несколько сложнее. Основная трудность состоит здесь, по-видимому, в следующем. В фонологии относительно легко проверить все позиции звука с тем, чтобы установить, является ли некоторый признак связанным или нет. Для слов практически невозможно проверить все позиции, чтобы установить, является ли категория морфологической (в смысле приведенного определения) или нет. Возьмем в качестве примера n а д е ж и. Кажалось бы, все падежи существительного, например, в русском языке, за исключением именительного, связаны позицией (скажем, управлением). Однако можно найти позиции, где, например, винительный падеж может быть заменен родительным («дай мне *вино*» — «дай мне *вина*»), творительный — именительным («он был *поэт*» — «он был *поэтом*», «*течение* отнесло лодку» — «*течением* отнесло лодку»), родительный — творительным («полный *мыслей*» — «полный *мыслями*»), творительный — винительным («он швырял в нее *камнями*» — «он швырял в нее *камни*»). Из чешского языка можно привести пример параллельного употребления винительного и дательного («*učiti se neso*» — «*učiti se neso*»).

Параллельное употребление этих падежей (для личных местоимений) зафиксировано и в болгарском языке («Бог да ги прости», ср. «Аз и прощавам», аналогично для глаголов *любвам се, радвам се, чудя се, сърдя се, вервам, завиждам*⁴⁰). Нельзя забывать и о свободном употреблении падежей, ср. многочисленные примеры более или менее свободного употребления русских падежей, приведенные в известной работе Р. Якобсона⁴¹.

Откажемся поэтому от построения полного изоморфизма между системой фонетических признаков и системой категорий. Будем считать все категории морфологическими (приблизившись таким образом к определению морфологической категории в указанной работе И. А. Мельчука).

Теперь грамматема соответствует элементарной грамматической категории. В дальнейшем мы увидим, что понятие категориальной однородности, т. е. отсутствия пустых клеток чрезвычайно важно при выборе той или иной системы категорий. При интерпретации соответствующих моделей мы будем в основном иметь дело с категориями имен. Дело в том, что для имени соответствующая подсистема может быть, правда иногда с некоторыми натяжками (см. § 34), сведена к категориально однородной. В самом деле, замена одного падежа другим или одного числа другим и т. п. в подавляющем большинстве случаев приводит к грамматеме, имеющейся в данном языке.

Для подсистемы глагола пустые клетки, пожалуй,— не исключение, а правило. Замена категории «изъявительное наклонение» на «повелительное» или «условное», как правило, приводит к потере категорий времени и т. п.

Может быть, это связано с тем, что категории имени обслуживают, как правило, только одну функцию языка, а именно функцию номинативную (мы отвлекаемся здесь и в дальнейшем от форм типа звательного падежа, представленного в ряде славянских языков), в то время как в категориях глагола пересекаются самые разнообразные функции, в первую очередь коммуникативные (ср. хотя бы анализ этих сложно пересекающихся функций в терминах теории информации Р. Якобсоном⁴²). Тем не менее можно надеяться, что ряд понятий модели, интерпретируемых на системе имени, могут оказаться полезными и для интерпретаций на системе глагола.

⁴⁰ См.: Е. В. Чешко. Падежи и предлоги в современном болгарском литературном языке. «Основные вопросы грамматики болгарского языка», М., 1960, стр. 25, 43.

⁴¹ R. Jakobson. Beitrag zur allgemeinen Kasuslehre.— TCLP, Prag, 1936, 6, стр. 240—287.

⁴² R. Jakobson. Shifters, verbal categories and the Russian verb, Harvard, 1957.

В этом параграфе мы показали, как можно использовать модель, перенесенную из фонологии, для моделирования плана содержания в грамматике. Однако мы не можем в качестве исходной гипотезы принять изоморфизм плана выражения и плана содержания в грамматике.

Для исследования парадигматических отношений в плане выражения, т. е. отношений между классами слов, выделяемых на основе анализа распределения (семейства), и семантической информацией, передаваемой словами (задание окрестностей), мы воспользуемся сначала моделями, в которых категории, т. е. единицы плана содержания в области грамматики, вообще не рассматриваются.

Если в модели, построенной для анализа единиц плана содержания, основную роль играло понятие категориальной однородности, то в моделях, которые мы рассмотрим в следующей главе, главным инструментом исследования будет понятие формальной однородности. Затем будет построена модель, учитывающая соответствия между планом выражения и планом содержания, т. е. модель, в которой строится отображение множества слов на множество категорий.

Примечания.

1) В нашей модели категории задаются извне. По-видимому, в дальнейшем целесообразно строить модели, в которых исследование формальных свойств категорий, проводимое в нашей работе, сочетается с анализом общих значений категорий и выделением их «дифференциальных признаков».

Мы видели, что задание грамматических категорий вполне аналогично заданию фонетических признаков. В § 10 мы говорили о системе дифференциальных признаков, предложенной Р. Якобсоном для фонологии и сводящей все разнообразие признаков к некоторому небольшому числу дифференциальных элементов. Подобная же идея была осуществлена Р. Якобсоном⁴³ для падежей. К сожалению, все эти исследования еще не доведены до такого состояния, чтобы их можно было уже сегодня применять к формальной модели. Исходя из этого соображения, а также из желания облегчить изложение и без того трудного материала, мы оперируем здесь традиционными категориями. По-видимому, все полученные при такой интерпретации результаты можно будет перенести и на случай задания не категорий, а дифференциальных признаков.

2) Идея дифференциальных признаков приложима, по-видимому, и к семантике. Некоторые интересные идеи были высказаны в этой связи Вяч. Вс. Ивановым⁴⁴. Они сводятся к следующему.

⁴³ Ср. также: Р. Якобсон. Морфологические наблюдения над славянским склонением. «American Contributions to the 4. International Congress of Slavists». 's-Gravenhague, 1958.

⁴⁴ Вяч. Вс. Иванов. Понятие нейтрализации в морфологии и лексике. — «Бюллетень объединения по проблемам машинного перевода». № 5, 1957, стр. 56.

Задается разбиение всего множества слов на «семантические поля», например поле слов, обозначающих деятеля, поле слов, обозначающих умственные способности людей и т. п. «Значение каждой лексической единицы можно рассматривать как множество, состоящее из счетного числа «сем» (или «семантических дифференциальных признаков»). Выбор «сем», из которых состоит значение данной лексической единицы, определяется числом семантических полей, в которое входит эта единица».

В связи с этим ряд определений из главы II могут быть, по-видимому, перенесены не только на грамматические категории, но и на «семантические поля». Однако соответствующие идеи пока еще только намечены.

ПАРАДИГМАТИЧЕСКИЕ МОДЕЛИ
В ГРАММАТИКЕ

§ 27. Моделирование понятия «часть речи»

Первая парадигматическая модель, которую мы рассмотрим, была построена О. С. Кулагиной в связи с осмыслением некоторых проблем, встающих при машинном переводе. Поскольку в грамматиках описание системы языка обычно ведется в терминах «частей речи», то модель имела в виду формализовать это весьма неточное понятие.

Как указано в § 24, само это понятие с формальной точки зрения имеет лишь ограниченную ценность. Однако построенная О. С. Кулагиной модель¹ дала возможность поставить ряд вопросов, выходящих далеко за рамки данной темы. В этом прежде всего лингвистическая ценность данной модели.

В модели Кулагиной рассматриваются всевозможные В-разбиения исходного множества слов и всевозможные В-структуры (§ 22). Для некоторого фиксированного разбиения В устанавливается В-эквивалентность двух множеств V_i и V_j .

В-эквивалентность, введенная в § 23, как легко проверить, обладает свойствами рефлексивности, симметрии и транзитивности. Тем самым при данном разбиении В все исходное множество слов разбивается на ряд непересекающихся подмножеств, таких, что любые два эквивалентные элемента V_i и V_k попадают в одно подмножество, а любые два неэквивалентные элемента V_i и V_j в разные подмножества. Это новое разбиение нашего множества на непересекающиеся классы, порожденное разбиением В, называется производным от разбиения В и обозначается V^* .

Итак, V^* — это такое разбиение исходного множества слов,

¹ О. С. Кулагина. Об одном способе определения грамматических понятий на базе теории множеств. «Проблемы кибернетики», вып. 1. М., 1958.

при котором в каждый класс этого разбиения (обозначим его V_i^*) попадают все эквивалентные между собой элементы разбиения V и только они.

Поскольку V^* -разбиение есть частный случай V -разбиения (как произвольного разбиения на непересекающиеся классы), то можно было бы считать, что подобным же образом можно перейти от V^* к V^{**} и т. д.

Однако имеет место следующая замечательная теорема, доказанная О. С. Кулагиной.

Теорема 4.1. Производное разбиение V^{**} от производного разбиения V^* совпадает с разбиением V^* .

Лингвистический смысл теоремы, по-видимому, в следующем. Если исходить из круга понятий, введенных в § 20—21, а именно: слово и отмеченная фраза, то нам в сущности задано только одно V -разбиение, разбиение на единичные слова (E -разбиение). Исходя из этого разбиения, мы получили понятие E -эквивалентности. Отсюда мы получили в сущности производное от E -разбиения, а именно разбиение на семейства (S -разбиение). Казалось бы, что теперь можно ввести понятие S -эквивалентности и получить новое V -разбиение и т. д. Теорема показывает, что это невозможно. Между тем разбиение на семейства — лишь простейшее из тех разбиений, с которыми мы имеем дело в языке, а E -эквивалентность лишь простейшая из таких эквивалентностей. Для того чтобы формализовать другие более сложные эквивалентности, заданных в § 20—21 понятий недостаточно. Нужно ввести еще хотя бы одно разбиение, заданное извне. Таким разбиением и является разбиение на окрестности, введенное в § 24. Мы видим, таким образом, что для введения понятия окрестности, которое мы мотивировали содержательными аргументами, имеются и важные формальные основания. В модели, которая использует кроме «слова» и «отмеченной фразы» еще и понятие «окрестность», мы можем вводить в рассмотрение новые V -эквивалентности и новые производные разбиения.

Разбиение на типы, или T -разбиение, определяется как производное от разбиения на окрестности, т. е. от T -разбиения: $T^*(x) = T(x)$.

Если теперь окрестность понимать в смысле интерпретации 1 (§ 24), то можно считать, что мы получили понятие, довольно близкое к «части речи»². В самом деле, возьмем такие окрестности: все формы прилагательного *большая, -ой, -ые*

² По крайней мере, если мы ограничимся словами в их «первичной» функции (ср. § 24). Ясно, что отглагольные существительные, сохраняющие глагольное управление, попадут в особые «части речи», отличные от «части речи», включающей слова *стол, окно, миска* и т. п., поскольку, например, во фразе «*овладение знаниями*» слово *овладение* нельзя заменить словом *окно*.

и т. п. и все формы прилагательного *сильный*. Ясно, что эти две окрестности эквивалентны между собой; вместе они объединяются в тип, куда входят как все слова, входящие в парадигму склонения слова *сильный*, так и все слова, входящие в парадигму слова *большой*, да и всех других прилагательных русского языка. Таким образом, Г (*большой*) очень близко к части речи «прилагательное». (Разумеется, этот тип не совпадает целиком с частью речи «прилагательное», так как не включает в себя прилагательные, способные к управлению, например *полный*, *богатый*, *уверенный* и т. п., но мы отвлекаемся от подобных случаев.) Пусть теперь все прилагательные объединены в один тип и пусть они являются единственными согласуемыми с существительными формами (от причастий, местоимений, глагольных форм прошедшего времени и т. д. мы также отвлекаемся). Покажем, что теперь в один тип попадают и существительные разных родов. Пусть у нас есть фразы:

*Маленький мальчик лежит на полу,
Старая кукла лежит на полу,
Золотое кольцо лежит на полу.*

Возьмем Г-структуры этих фраз:

Г(*маленький*) Г(*мальчик*) Г(*лежит*) Г(*на*) Г(*полу*),
Г(*старая*) Г(*кукла*) Г(*лежит*) Г(*на*) Г(*полу*),
Г(*золотое*) Г(*кольцо*) Г(*лежит*) Г(*на*) Г(*полу*).

Ясно, что Г(*маленький*) \sim Г(*старая*) \sim Г(*золотое*), поскольку, например, в Г(*маленький*) входят слова *маленькая* и *маленькое* и т. п. Отсюда ясно, что отмеченность Г-структуры:

Г(*маленький*) Г(*мальчик*) Г(*лежит*) Г(*на*) Г(*полу*)

не изменится от того, подставим ли мы вместо окрестности Г(*мальчик*) окрестности Г(*кукла*) или Г(*кольцо*), причем это верно и для произвольных Г-структур вида:

Г(А) Г(*мальчик*) Г(В).

Таким образом, все существительные (во всяком случае такие, как *мальчик*, *кукла*, *окно* и т. п.) также объединяются в один тип.

Оказалось, однако, что понятие типа нуждается в уточнении с формальной точки зрения. Дело в том, что, как показал В. А. Успенский, из принадлежности двух слов одному семейству вовсе не следует принадлежности их к одному типу. (В приложении приводится пример искусственного языка, в кото-

ром два слова из одного семейства относятся к разным типам.) В связи с этим были введены новые определения.

Назовем язык *правильным*³, если всякие два слова, принадлежащие к одному семейству, принадлежат и к одному типу. Назовем два элемента x и y R -эквивалентными, если существует цепочка элементов r_1, r_2, \dots, r_n , такая, что:

1) при всяком i элементы r_i и r_{i+1} принадлежат либо одной и той же окрестности, либо одному и тому же семейству;

2) $r_1 = x, r_n = y$.

Множество исходных элементов распадается (как легко проверить) на классы R -эквивалентных элементов, которые мы будем называть разделами.

Примером слов, принадлежащих к одному разделу, являются *столом* и *стул*. В самом деле между *столом* и *стул* можно построить такую цепочку: *стул* входит в то же семейство, что и *стол*, а *стол* входит в ту же окрестность, что и *столом*.

Имеют место следующие соотношения, сформулированные и доказанные В. А. Успенским:

Теорема 4.2. В правильных языках для всякого слова x верно следующее:

1) $R(x) \subseteq T(x)$, т. е. раздел содержится внутри типа;

2) $T(x) = R^*(x)$, т. е. разбиение на типы совпадает с производным от разбиения на разделы.

Таким образом, мы видим, что при определенной интерпретации понятия «окрестность» мы можем получить аналог понятия «части речи», а именно как производное от понятия «раздел» (разумеется, что здесь имеют место те же ограничения, что и раньше, а именно слова рассматриваются в их «первичной» функции).

Еще одно понятие, дающее приближение к части речи, было введено О. С. Кулагиной. Назовем *морфологическим классом* $N(x)$ слова x множество всех слов x' , таких, что выполнено хотя бы одно из условий:

1) $\Gamma(x) \cap S(x') \neq 0$,

2) $\Gamma(x') \cap S(x) \neq 0$.

Иными словами, в один морфологический класс объединяются такие слова, что в семействе одного из них имеется слово из окрестности другого. В языках, где морфологические классы образуют разбиение (например, французском), производное разбиение от разбиения на классы также довольно близко к части речи (о связи между этими разбиениями см. указанную работу О. С. Кулагиной, а также § 29).

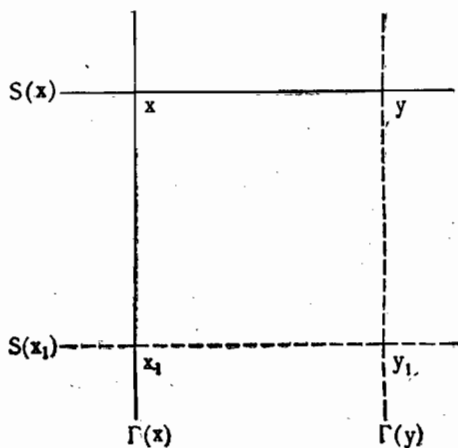
³ Это и дальнейшие определения и утверждения этого параграфа взяты из работы В. А. Успенского «К определению части речи в теоретико-множественной системе языка». — «Бюллетень Объединения по проблемам машинного перевода», № 5. М., 1957.

§ 28. Простота и однородность языка

В модели О. С. Кулагиной введено одно понятие, чрезвычайно важное для всего дальнейшего изложения. Она называет язык простым, если выполняются следующие два требования:

1) для любого x верно $\Gamma(x) \cap S(x) = x$,
иначе говоря, требуется, чтобы два слова, входящие в одну окрестность, принадлежали разным семействам;

2) если $x' \in \Gamma(x)$ и $y \in S(x)$, то $S(x') \cap \Gamma(y) \neq 0$,
иначе говоря, если имеется слово x_1 , входящее в $\Gamma(x)$, и x входит в $S(y)$, то должно существовать некоторое слово y_1 , входящее в $S(x_1)$ и $\Gamma(y)$. На чертеже это можно схематически представить следующим образом. Пусть линиями обозначены $S(x)$ и $\Gamma(x)$, а пунктиром $S(x_1)$ и $\Gamma(y)$. Тогда пунктирные линии должны пересекаться в некоторой точке y_1 .



Поскольку каждое из этих требований имеет определенный лингвистический смысл, то мы дадим им специальные названия, а именно первое требование будем называть требованием простоты языка, а второе — требованием формальной однородности языка. Второе понятие играет столь существенную роль в последующем изложении и будет употребляться столь часто, что в контекстах, где это понятие нельзя смешать с категориальной однородностью, мы будем просто говорить об однородности языка, опуская прилагательное «формальный».

Требование простоты, по-видимому, выполняется в русском языке и других славянских языках. Правда, в ряде языков встречаются пары типа *профессора* и *профессоры*, что нарушает требование простоты.

В связи с этим полезно ввести следующие понятия. Мы будем говорить об устранимой непростоте в случаях, когда не существует двух разных грамматических категорий K_1 и K_2 , таких, чтобы из двух слов x_1 и x_2 , входящих в одну окрестность и одно семейство, x_1 относилось бы к K_1 и не относилось к K_2 , а x_2 относилось бы к K_2 и не относилось бы к K_1 . В противном случае мы будем говорить о неустраимой непростоте.

Примечание. Если задание категорий в модели зависит лишь от свойств семейств и окрестностей, то понятие устранимой и неустраимой простоты внутри модели смысла не имеет. Однако оно очень удобно для интерпретации модели. Так, очевидно, что в приведенном примере из русского языка мы имеем дело с устранимой непростотой, так как ни в одной грамматической системе слова *профессоры* и *профессора* не будут, по-видимому, отнесены к каким-то двум разным категориям.

С принципиально неустраимой непростотой мы встречаемся при анализе финно-угорских языков, например финского или эстонского. Так, в эстонском языке формы комитатива (*Kaasaütlev*) входят в одно семейство с формами абессива (*Ilmaütlev*); т. е., например, не нарушая грамматической правильности фразы, можно заменить комитатив *rõõmutiga* на абессив *rõõmututa* и наоборот. Ср.: *Suure rõõmutiga tulen sinna* 'С большой радостью приду туда'; *Suure rõõmututa tulen sinna* 'Без большой радости приду туда'. Аналогичное положение с так называемыми внутренними и внешними падежами. Так, инессив *majas* 'в доме' входит в одно семейство с адессивом *majal*, элатив *majast* 'из дому' — в одно семейство с аблативом *majalt*, а иллатив *majja* 'в дом' — в одно семейство с аллативом *majale* 'к дому'. Случаи неустраимой непростоты интересны тем, что они создают непреодолимую преграду для строго дистрибутивного анализа, т. е. дистрибутивный анализ предполагает простоту языка.

С типологической точки зрения положение в падежной системе финно-угорских языков можно в какой-то мере сравнить с эквивалентностью разных глагольных форм русского языка, например *стучал*, *стукнул*, а также *стучит*, *стучал* и, очевидно, *стучал бы* (если считать это одной словоформой) относятся к парадигме *стучать*. Между тем эти формы представляются эквивалентными. Ср. Вчера он *стучит* ко мне — *praesens historicum*; Вчера он *стучал* ко мне; Вчера он *стучал бы* ко мне.

Из этих примеров видно, что *стучит* можно заменить на *стучал* и обратно. Интересно вспомнить в этой связи, что А. М. Пешковский считал видовые формы (типа *кольнул* — *колол*) «несинтаксическими» и одновременно склонялся к отнесению форм времени и склонения в число «несинтаксических»⁴.

⁴ А. М. Пешковский. Русский синтаксис в научном освещении. М., 1938, стр. 60.

Правда, в русском языке существует изменение глагола в прошедшем времени по родам, что мешает объединению *стучал* и *стучала* в одном семействе (на формы вида это, разумеется, не распространяется). Но, например, в немецком языке *sagte* и *sagt*, *spräche*, *sprach* и *spricht* входят в одно семейство и одну парадигму, т. е. немецкий язык не является простым.

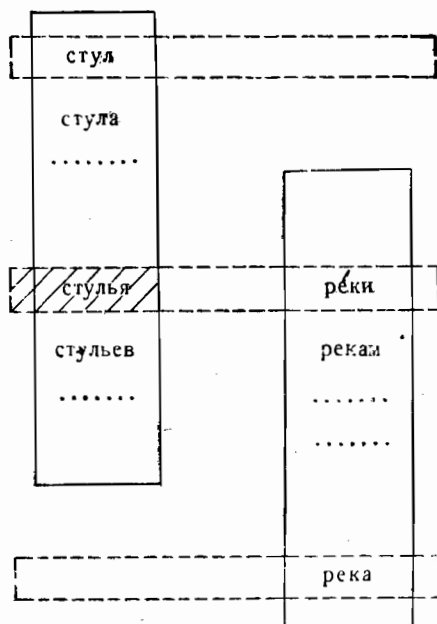
Второе требование, а именно требование однородности, еще более важно, чем первое, хотя, как будет видно из дальнейшего, реально однородных языков, по-видимому, не существует.

Сразу же выяснилось, что русский язык неоднороден. В самом деле:

$\Gamma(\text{стул}) \cap S(\text{рэки}) = \text{стулья}$ (слово *стулья* входит в окрестность слова *стул* и в одно семейство со словом *реки*).

$S(\text{стул}) \cap \Gamma(\text{рэки}) = 0$ (нет слова, входящего в окрестность слова *рэки*, которое входило бы в семейство, содержащее слово *стул*).

На чертеже это можно представить следующим образом:



Сплошной линией обведены окрестности, а пунктиром — семейства. Заштриховано пересечение семейства $S(\text{рэки})$ с окрестностью $\Gamma(\text{стуль})$.

Раньше многие (в том числе и автор⁵) считали, что пример однородного языка являет французский. Дело в том, что одной из причин неоднородности, например, в русском языке, является «склеивание» семейств во множественном числе. Во французском языке такого склеивания не происходит. Но во французском языке существуют *pluralia* и *singularia tantum*, т. е. слова, употребляющиеся только во множественном или только в единственном числе, например *moers* 'правы'. Поэтому:

$$S(\textit{moers}) \cap \Gamma(\textit{vertu}) = \textit{vertus} \text{ 'добродетели'},$$

$$\Gamma(\textit{moers}) \cap S(\textit{vertu}) = 0.$$

Понятие однородности полезно прежде всего как модель, как некоторая абстракция.

В частности, понятие формальной однородности следующим образом связано с ранее введенным понятием правильности языка.

Теорема 4.3 (Успенский). Всякий формально однородный язык есть правильный язык.

Еще более важно понятие приближения к однородности, хотя это понятие пока еще формализовать не удалось. Мы покажем лишь, что при анализе реальных языков и при сравнении их между собой можно, по-видимому, ставить вопрос о степени однородности языка. Так, можно считать, что польский язык еще менее однороден, чем русский. Дело в том, что в польском языке в единственном числе различаются три ряда, как и в русском, что дает три разных семейства, например:

- 1) *chłopiec, poeta, śnieg* ('парень', 'поэт', 'снег');
- 2) *córka, żona* ('дочка', 'жена');
- 3) *dziecko* ('дитя').

Во множественном числе в польском языке, однако, различаются два рода (в русском же все слова множественного числа принадлежат грамматически одному роду), а именно мужской личный и немужской, включая сюда все слова мужского рода, обозначающие предметы. Различия проявляются в форме согласуемых с данными слов, например:

$$\textit{dobrzy} \text{ 'хорошие'} \begin{cases} \textit{chłopcy}; \\ \textit{poety}; \end{cases} \quad \text{но } \textit{dobre} \begin{cases} \textit{dziecy} \\ \textit{córky} \\ \textit{żony} \\ \textit{śniegi} \end{cases}$$

или

$$\textit{czytali} \text{ 'читали'} \begin{cases} \textit{chłopcy}; \\ \textit{poety}; \end{cases} \quad \text{но } \textit{czytały} \begin{cases} \textit{dzieci} \\ \textit{córky} \\ \textit{żony} \end{cases}$$

⁵ См.: И. И. Ревзин. О некоторых понятиях теоретико-множественной концепции языка.— ВЯ, 1960, № 6, стр. 93.

Таким образом, трем разным семействам в единственном числе (словам мужского неличного, женского и среднего рода) соответствует одно семейство множественного числа. Это положение аналогично положению в русском языке, где трем семействам (мужского, среднего и женского рода) соответствует одно семейство во множественном. Но в польском языке, кроме того, словам, принадлежащим двум разным семействам во множественном числе (а именно мужского личного рода и мужского неличного рода), соответствует одно семейство в единственном числе.

С другой стороны, чешский язык явно более однороден, чем польский, и даже чем русский. В чешском языке различаются три семейства в единственном числе, например:

М. р.	{	<i>nový žák</i>	‘новый ученик’
		<i>nový sešit</i>	‘новая тетрадь’
Ж. р.		<i>nová tužka</i>	‘новый карандаш’
Ср. р.		<i>nové pero</i>	‘новое перо’

и три семейства во множественном числе, устроенные иначе, чем в единственном числе:

М. р. личн.	<i>novi žáci</i>
М. р. неличн.	{ <i>nové sešity</i>
Ж. р.	{ <i>nové tužky</i>
Ср. р.	<i>nová pera</i>

Таким образом, в чешском языке слова, относящиеся к двум семействам единственного числа, входят в одно семейство множественного и слова, относящиеся к двум семействам множественного, входят в одно семейство единственного. Сопоставим все три языка — русский, польский и чешский в соответствующей таблице:

Язык	Число	М. р. личн.	М. р. неличн.	Ж. р.	Ср. р.
Польский	ед.	_____	_____	_____	_____
	мн.	_____	_____	_____	_____
Чешский	ед.	_____	_____	_____	_____
	мн.	_____	_____	_____	_____
Русский	ед.	_____	_____	_____	_____
	мн.	_____	_____	_____	_____

С типологической точки зрения интересно, что все эти языки можно было бы сравнивать, исходя из того, какова степень устранимости неоднородности в каждом из них.

Так, например, в чешском языке легко можно добиться однородности; если исключить из рассмотрения все предметные (неодушевленные) существительные мужского рода (как *sešit* в нашем примере), то, как легко проверить, в оставшейся части чешский язык будет однородным. В польском языке, где во множественном числе различаются всего два рода, которые к тому же не совпадают с выделяемыми в единственном числе, или в русском языке, где во множественном числе выделяется только один род, неоднородность устранить гораздо трудней.

Из славянских языков, пожалуй, наиболее близок к однородному языку (если отвлечься от болгарского) сербско-хорватский, в котором — для рассматриваемого нами фрагмента — каждое семейство в единственном числе соответствует ровно одному семейству во множественном числе, и обратно, ср. *mlád jélen* — *mládi jélени*, *mládo дрво* — *mláda дрвета* и *mláda жена* — *mláde жéне*.

Мы видели, что степень устранимости неоднородности может явиться полезным типологическим критерием при сравнении славянских языков. Он интересен и для сравнения славянских языков с неславянскими, входящими в тот же языковой союз, что и некоторые славянские языки. Так, в румынском языке ряд исследователей выделяет группу слов так называемого личного рода, которая сравнивается с лично-мужским родом польского языка. Как показал недавно И. К. Кицимия, с типологической точки зрения мы здесь имеем дело в этих двух языках с совершенно разными явлениями⁶. К аргументам, приведенным этим автором, можно добавить следующий: наличие так называемых личных имен в румынском языке не приводит к явлениям неоднородности, что так характерно для лично-мужского рода польского языка.

Интересно, однако, что в румынском языке есть явления, типологически сравнимые с неоднородностью польского, чешского или словацкого языков. Речь идет, однако, не о «личном роде», а о так называемом среднем или обоюдном роде румынского языка, а именно о наличии ряда слов, которые принадлежат к мужскому роду в единственном числе и к женскому роду во множественном числе⁷; таким образом, семейства, получаемые в единственном числе, не совпадают с семействами во множественном числе.

⁶ И. К. Кицимия. Личный род в польском и румынском языках. «Revue de linguistique». Bucarest, 1960, № 1.

⁷ См.: К. Pátruș. Sur le genre «neutre» en roumain. «Mélanges linguistiques publiée à l'occasion du VII-e congrès International des linguistes». Bucarest, 1957

§ 29. Подокрестность и подсемейство.
Связь между подсемейством и элементарной
категорией

Мы будем говорить, что два слова x и y , принадлежащие одной окрестности, входят в одну подокрестность (обозначим подокрестность слова x через $\Pi^1(x)$), если в окрестности любого слова, входящего в то же семейство, что и x , найдется слово w , входящее в то же семейство, что и y . Так, в русском языке слово *дом* входит в ту же подокрестность, что и слово *дому*, ибо они, во-первых, входят в одну окрестность, а во-вторых, в окрестности любого слова, входящего в то же семейство, что и слово *дом* (например, возьмем окрестность слова *стол*) имеется слово, входящее в то же семейство, что и слово *дому* (в нашем примере таким словом является *столу*). В то же время слова *дом* и *домов*, входящие в одну окрестность, входят в разные подокрестности, ибо имеется слово *окон*, входящее в то же семейство, что и слово *домов*, но в окрестности слова *окон* нет слова, входящего в то же семейство, что и слово *дом*.

Аналогичным способом введем теперь понятие подсемейства. Мы будем говорить, что два слова x и y , принадлежащие одному семейству, входят в одно подсемейство (подсемейство слова x обозначим $\Pi S(x)$), если в семействе любого слова z , входящего в ту же окрестность, что и x , найдется слово w , входящее в ту же окрестность, что и y . Так, в русском языке слово *дом* входит в то же подсемейство, что и слово *стол*, ибо они, во-первых, входят в одно и то же семейство, а во-вторых, в семействе любого слова, входящего в ту же окрестность, что и *дом* (например, возьмем семейство слова *дому*), имеется слово, входящее в ту же окрестность, что и *стол* (в нашем примере таким словом опять будет слово *столу*).

Примеры показывают, что выделение подокрестностей связано с выделением подсемейств. Но этого и следовало ожидать, поскольку определения подокрестности и подсемейства дословно повторяют друг друга, лишь «окрестность» заменяется всюду на «семейство», а «семейство» на «окрестность».

Легко видеть, что если для всякого x подокрестность совпадает с окрестностью или для всякого x подсемейство совпадает с семейством, то язык будет формально однородным.

Однако понятия подокрестности и подсемейства введены не для того, чтобы получить новую формулировку понятия однородного языка, а для анализа неоднородных языков.

В § 28 мы рассматривали некоторый фрагмент польского, чешского и русского языков. Там мы зафиксировали некоторый падеж (именительный) и смотрели, как распределяются в терминах семейств и окрестностей формы числа и рода.

Если же теперь проанализировать распределение подсемейств в этом фрагменте, то получим следующую картину:

Язык	Число	М. р. личн.	М. р. неличн.	Ж. р.	Ср. р.
Польский	ед.	████████	████████	████████	████████
	мн.	████████	████████	████████	████████
Чешский	ед.	████████	████████	████████	████████
	мн.	████████	████████	████████	████████
Русский	ед.	████████	████████	████████	████████
	мн.	████████	████████	████████	████████

Приведенная таблица интересна со следующей точки зрения. Хотя в анализированных нами фрагментах польского и чешского языков выделяются различные семейства и они ведут себя различно с точки зрения устранимости неоднородности, мы получаем в них те же самые подсемейства (т. е. существует изоморфное соответствие между системой подсемейств чешского языка и системой подсемейств польского языка).

Рассмотрим теперь более широкий фрагмент русского языка, включая и изменения по падежам. Пусть нам заданы следующие окрестности:

<i>корова</i>	<i>лампа</i>	<i>стол</i>	<i>слон</i>	<i>окно</i>	<i>тело</i>
<i>корове</i>	<i>лампе</i>	<i>стола</i>	<i>слона</i>	<i>окна́</i>	<i>те́ла</i>
<i>корову</i>	<i>лампу</i>	<i>столу</i>	<i>слону</i>	<i>окну</i>	<i>телу</i>
<i>коровам</i>	<i>лампам</i>	<i>столам</i>	<i>слонам</i>	<i>окнам</i>	<i>телам</i>
<i>коров</i>	<i>ламп</i>	<i>столов</i>	<i>слонов</i>	<i>окон</i>	<i>тел</i>
<i>коровы</i>	<i>лампы</i>	<i>столы</i>	<i>слоны</i>	<i>о́кна</i>	<i>тела́</i>

Анализируя отмеченные фразы, в которых эти слова встречаются, мы получим следующие семейства, а внутри них, пользуясь определениями этого параграфа, следующие подсемейства (которые выделены пунктирной рамкой):

1. корова лампа
2. стол

эквивалентного в окрестности слова *столу* (слова *стол* и *слон* не эквивалентны, ср. «я вижу *стол*» при невозможности сказать «я вижу *слон*»). Аналогичное рассуждение можно провести и для слов *корова* и *лампа*.

Значение приведенного примера следующее. Если вспомнить понятие элементарной грамматической категории, введенное в § 25, то можно установить, что в наших примерах два слова x и y , принадлежащие к одному подсемейству, соответствуют той же самой или тем же самым элементарным грамматическим категориям. В связи с этим введем следующее определение, связывающее понятия подсемейства и элементарной грамматической категории. Возьмем некоторый язык, заданный как множество отмеченных фраз, с разбиением всех слов на окрестности и категории. Мы будем называть язык *регулярным*, если в нем любые два слова x и y , принадлежащие к одному подсемейству, относятся к той же самой элементарной категории (случай, когда подсемейство соответствует дизъюнкции элементарных категорий, т. е. когда мы имеем дело с грамматической омонимией, будет рассмотрен в § 34).

Теперь мы хотим поставить вопрос, являющийся одним из кардинальных вопросов всякого грамматического описания, а именно: сколько разных грамматических категорий целесообразно задать для того, чтобы это задание категорий соответствовало формальным свойствам языка. Исходная ситуация такова. Нам задан некоторый регулярный язык, для которого получено разбиение на подсемейства. Кроме того, имеется множество элементарных категорий, каждая из которых представляет собой пересечение нескольких неоднородных категорий. Нам задано также отображение множества всех слов на множество элементарных категорий. При этом отображении в силу регулярности языка каждый полный прообраз некоторой элементарной категории, т. е. множество слов, поставленных в соответствие с данной элементарной категорией, совпадает с каким-то подсемейством.

Возьмем теперь некоторое слово y . Назовем морфологическим подклассом слова y , который обозначим $M(y)$, совокупность всех слов x , таких, что пересечение подсемейства, содержащего x , и окрестности, содержащей y , не пусто. Образует для каждого слова его морфологический подкласс. Оказывается, что верно следующее утверждение.

Теорема 4.4: а) Морфологические подклассы образуют разбиение всего множества слов; б) производное от разбиения на морфологические подклассы совпадает с разбиением на типы.

С л е д с т в и е. В случае формально однородного языка эти утверждения верны для морфологических классов⁸.

⁸ Ср. теоремы 2 и 3 в цитированной работе О. С. Кулагиной.

Можно было бы ввести аналогичным способом понятие псевдоморфологического подкласса как совокупности слов, таких, что в ней для любой пары слов x и y пересечение подокрестности, содержащей x , с семейством, содержащим y , не пусто. Было бы легко показать, что для подобных подклассов можно сделать утверждение, аналогичное теореме 4.4а. Но, по-видимому, псевдоморфологические подклассы не имеют лингвистического смысла, в то время как относительно морфологических подклассов имеет место следующее важное утверждение:

Теорема 4.5. Если язык регулярен, то два слова, относящиеся к одному морфологическому подклассу, входят в одни и те же парадигматические категории.

Посмотрим теперь, какие морфологические подклассы участвуют в примере, рассмотренном выше. Слова *окну* и *телам* входят в один морфологический подкласс, поскольку имеется слово *телу*, входящее в то же подсемейство, что и слово *окну*, и в ту же окрестность, что и слово *телам*. Легко убедиться в том, что к этому морфологическому подклассу относятся все слова среднего рода. К другому морфологическому подклассу относятся все одушевленные существительные мужского рода: *слон*, *человеку*, *мужам* и т. п. К третьему морфологическому подклассу относятся все неодушевленные существительные мужского рода: *стол*, *стулья* и т. п. К четвертому морфологическому подклассу относятся все одушевленные существительные женского рода: *коровам*, *жене*, *дама* и т. п. Наконец, к пятому морфологическому подклассу относятся все неодушевленные существительные женского рода: *лампу*, *бритвам* и т. п. Наш пример показывает, что в русском языке не только два слова, входящие в один морфологический подкласс, относятся к одним и тем же парадигматическим категориям, но и два слова, входящие в разные морфологические подклассы, отличаются друг от друга хотя бы одной парадигматической категорией. Иначе говоря, в русском языке можно установить, по-видимому, взаимно однозначное соответствие между числом морфологических подклассов и числом конъюнкций парадигматических категорий, которые ставятся в соответствие со словами некоторого типа. Этот принцип мы и примем в модели. Он удобен тем, что число морфологических подклассов определяется независимо от задания категорий (оно зависит лишь от задания окрестностей и множества отмеченных фраз).

Разумеется, число морфологических подклассов еще не определяет однозначно числа парадигматических категорий, ибо разные неоднородные категории могут пересекаться, что оставляет возможность разным способом задавать категории, соблюдая при этом условие регулярности. Например, мы получили пять морфологических подклассов. Задать категории мы можем

по меньшей мере двумя разными способами. Во-первых, можно задать, как это делается обычно, две группы категорий:

- 1) мужской род, женский род, средний род;
- 2) неодушевленность, одушевленность.

Заметим, однако, что такое задание категорий, при наличии пяти (а не шести) морфологических подклассов, приводит к категориальной неоднородности, поскольку в русском языке категории одушевленности и неодушевленности ставятся в соответствие лишь со словами мужского и женского рода (аналогичное положение, как видно из приведенных ранее примеров, и в других славянских языках). Можно поэтому поступить иначе и задать пять однородных парадигматических категорий: одушевленно-мужской род, неодушевленно-мужской род, одушевленно-женский род, неодушевленно-женский род, средний род.

Мы бы предпочли именно такое задание категорий (поскольку оно удовлетворяет требованию категориальной однородности), однако формальный анализ средствами нашей модели позволяет лишь косвенно оценить общее число парадигматических категорий и не дает возможности определить конкретное задание категорий. Последнее может быть получено только в результате анализа дифференциальных признаков категорий в плане содержания, а это выходит за рамки данной работы (ср. первое примечание в § 26).

Обратимся теперь к заданию непарадигматических категорий. Интересен тот эмпирический факт, что во всех рассмотренных ранее примерах два слова, относящиеся к одному семейству, как правило, входят в одни и те же непарадигматические категории. Исключение составляют лишь случаи из эстонского языка, приведенные в § 28. Поскольку, однако, в славянских языках простота устранима, то мы от подобных случаев отвлечемся и введем следующее определение. Будем называть язык *н о р м а л ь н ы м*, если в нем каждое семейство, пересекающееся с некоторой непарадигматической категорией, целиком входит в нее. Для такого языка число семейств и будет в пределах данного типа определять число конъюнкций непарадигматических категорий (это вполне соответствует условию, выдвинутому в § 25). Что касается задания отдельных рядов однородных категорий, то здесь действительно все, что сказано относительно парадигматических категорий.

Посмотрим теперь, как связаны между собой столь важные требования, как формальная однородность и нормальность. Мы помним, что в формально однородном языке подсемейство совпадает с семейством. Поэтому, если язык формально однороден и регулярен, то он нормален, поскольку все слова, входящие в данное семейство, соответствуют заданному пересечению категорий, и, стало быть, каждой категории. Можно ввести

еще одно понятие, позволяющее связать задание категорий с формальной однородностью языка. Назовем непарадигматическую категорию равномерно распределенной, если для любого морфологического класса верно одно из двух: пересечение данной категории с любой окрестностью внутри него пусто или пересечение данной категории с любой окрестностью внутри него не пусто.

Примерами равномерно распределенных категорий являются все падежи русского языка. Однако не в каждом языке любой падеж есть равномерно распределенная категория. Например, в армянском языке местный падеж (на *-um*) употребляется только для существительных, обозначающих неодушевленные предметы (и даже не для всех существительных этой категории).

Неравномерно распределенной категорией является число в русском языке, да и в любом другом языке, где имеются так называемые *singularia* и *pluralia tantum*.

Имеет место следующее чрезвычайно важное утверждение:

Теорема 4.6. Если язык формально однороден и регулярен, то в нем непарадигматические категории распределены равномерно.

Это утверждение важно потому, что наличие неравномерно распределенных категорий делает формально неоднородным любой язык, для которого выполняется требование регулярности. Как мы уже указывали, ранее ошибочно считалось, что французский язык формально однороден (ср. § 28). Между тем достаточно было бы указать на то, что категории числа в этом языке распределены неравномерно, чтобы иметь возможность построить пример, опровергающий предположение о формальной однородности французского языка.

§ 30. Вопрос о разграничении индивидуальных парадигм

Мы знаем, что окрестности задаются, вообще говоря, произвольно. До сих пор мы исходили из интерпретации 1 (§ 24). Такое решение вопроса существенно упрощало все рассуждения, и поэтому было необходимо для обсуждения абстрактной модели.

Однако, как мы выяснили в § 24, при лингвистической интерпретации модели нам хотелось бы, чтобы задание окрестности отвечало двум требованиям:

1) все слова, входящие в одну окрестность, должны нести одинаковую семантическую информацию или — поскольку данное понятие не является достаточно ясным — все слова, входящие в одну окрестность, должны иметь одну и ту же лексическую морфему;

2) объединение слов в окрестности должно отражать некоторые системные отношения в языке или — поскольку данное понятие не является ясным — должно быть произведено таким образом, чтобы уменьшить некоторым нетривиальным способом степень неоднородности языка.

Поясним эти требования. Ясно, что объединение слов с разными лексическими морфемами противоречит лингвистической интерпретации понятия «окрестность» как системы форм данного слова. Единственное, что здесь можно допустить, это объединение так называемых суплетивных форм, т. е. алломорфов одной и той же лексической морфемы (например, *есть* и *быть*, *хорошо* и *лучше* и т. п.). Итак, слова одной окрестности должны выделять с точностью до алломорфа одну и ту же лексическую морфему. Однако это необходимое требование не является достаточным. В самом деле, нецелесообразно, с точки зрения лингвистической интерпретации, включать в одну окрестность любую пару из ряда слов: *человек*, *человечный*, *человечек*, *человеческий*, *человечество*, *человечина*, *человечище*, хотя в них, с точностью до алломорфа, одна и та же лексическая морфема *человек*. Конечно, интуитивно разница между этими словами и невозможность объединения их в одну парадигму ясна, хотя бы потому, что многие из них принадлежат к разным частям речи. Но «часть речи» мы определяем на основании определенной интерпретации понятия «окрестность», потому формального критерия для объединения в окрестности мы здесь не получаем.

Возникают проблемы, которые неоднозначно решаются и лингвистами. Должны ли мы считать, что *видеть* и *увидеть*, т. е. две видовые формы, принадлежат разным парадигмам или одной? В традиционных терминах эту проблему можно поставить следующим образом: относится ли образование видовых форм к с л о в о о б р а з о в а н и ю (две разные окрестности) или к ф о р м о о б р а з о в а н и ю (одна и та же окрестность). Как известно, этот вопрос по-разному решается разными лингвистами⁹. Ясно, что решение его должно быть найдено не на основании внешних критериев, а исходя из системы языка, из отношений, существующих внутри нее.

В наших терминах системность языка, симметричность, наблюдаемая в нем, отражается как степень однородности языка. Отсюда ясно, что слова надо так объединять в окрестности, чтобы язык становился более однородным (и, может быть, более простым).

Поэтому можно себе представить дело так, что сначала слова объединяются в окрестности на основе тождества лексиче-

⁹ См.: В. В. Виноградов. Современный русский язык, вып. II. М., 1938, стр. 380—390.

ских морфем (интерпретация 3), затем полученный таким образом язык исследуется с точки зрения его однородности. Если степень неоднородности получившегося языка высока, то делается попытка сократить окрестности таким образом, чтобы повысить степень однородности языка.

Если это удастся, то новое разбиение на окрестности и считается более правильным для данного языка.

Вопрос этот, оказывается, однако, более сложным, чем может показаться на первый взгляд. Дело в том, что совершенно тривиальным способом каждый язык можно превратить в однородный, а именно оставить в каждой окрестности ровно по одному слову. Такое решение может, однако, удовлетворить лингвиста лишь в том случае, если речь идет об изолирующем («аморфном») языке типа китайского, где действительно каждая окрестность содержит ровно одно слово (между прочим, в наших терминах может быть дано следующее определение «аморфного» языка: язык называется аморфным, если в нем для любого x верно $\Gamma(x) = \{x\}$).

Другим, несколько менее тривиальным способом является разделение первоначальных окрестностей на подокрестности и задание новых окрестностей, равных этим подокрестностям. Как мы знаем, в этом случае язык становится однородным.

Если, как в примерах § 28, ограничиться существительным, то, как мы видели, неоднородность ряда славянских языков объясняется в значительной степени («смещением») во множественном числе форм категорий рода. Если теперь формы единственного числа объединить в одной окрестности, а формы множественного числа в другой, то степень однородности языка существенно повысится. Между прочим, точка зрения, по которой образование множественного числа есть образование нового слова, а не формы этого же слова, неоднократно высказывалась в лингвистике; например, П. С. Кузнецов пишет: «Следуя Ф. Ф. Фортунатову, считаю формы единственного и множественного числа того, что обычно считается одним словом, разными словами, т. е. категориями, которые связаны между собой словообразовательными, а не словоизменительными отношениями»¹⁰. Если встать на эту точку зрения, то степень однородности русского и других славянских языков значительно повысится, и между ними вряд ли проявятся те существенные различия, о которых говорилось в § 28. Однако именно для славянских языков, как показал сам П. С. Кузнецов¹¹, характерен последовательный параллелизм падежных форм единственного и множественного числа. Поэтому

¹⁰ П. С. Кузнецов. Развитие индоевропейского склонения в общеславянском языке. М., 1958, стр. 7.

¹¹ Там же, стр. 49.

представление их как форм одного слова вполне оправдывается внутренними соображениями системы.

По-видимому, целесообразна такая постановка вопроса, при которой учитывается не только степень приближения к однородному языку, но и то, ценой каких усилий это достигается. Уточним эту сторону вопроса.

Во-первых, важно, чтобы отношение числа новых окрестностей к числу первоначальных окрестностей не было слишком большим. Именно это интуитивное соображение заставляет нас отказаться от деления каждой окрестности существительного на две (единственное число и множественное число).

Во-вторых, важно исключить некоторые слова, которые вызывают неоднородность. Например, наличие некоторых слов *pluralia* и *singularia tantum* во французском языке делает этот язык неоднородным. Возможны в принципе два способа устранения этой неоднородности: рассмотренное выше разделение на две окрестности форм единственного и форм множественного числа и удаление из рассмотрения слов *pluralia* и *singularia tantum*. Так как указанных слов мало, то первое решение гораздо более приемлемо, чем деление всех окрестностей, производимое именно из-за наличия данных слов.

Итак, наша задача может быть формализована следующим образом. Задано некоторое разбиение на семейства и некоторое разбиение на окрестности. Нужно наиболее экономным способом повысить степень однородности языка, причем разрешено исключить не более определенной доли от общего числа окрестностей и разрешено делить окрестности на части так, чтобы отношение числа новых окрестностей к общему числу окрестностей не превышало некоторое наперед заданное число.

Интересно вообще, в каких случаях подобная задача имеет решение. Пока еще все эти вопросы не сформулированы достаточно четко, чтобы допустить математическое решение, но ясна важность самой постановки подобной проблемы.

Интересна и другая задача, близкая к данной. Представим себе язык как систему, устойчивость которой обеспечивается достаточно высокой степенью однородности. Эта система испытывает некоторые возмущения¹², состоящие:

- 1) в появлении новых слов, не имеющих тех форм, которые имеют остальные слова данного типа (например, появление слов типа *пальто*, *кофе*, *метро* в русском языке);
- 2) в совпадении разных форм в результате некоторых фонетических процессов (например, совпадение в общеславянском форм родительного и дательного падежей единственного числа в основах на *-i*);

¹² О понятиях «устойчивости» и «возмущения» см.: У. Р о с с е й ш б и. Введение в кибернетику. Перевод с англ. М., 1959, глава V.

3) в появлении новых форм (например, формы на *-у* для слов типа *воску, чаю* и т. п.).

Все эти возмущения сводятся к тому, что появляются некоторые новые окрестности или изменяется количество форм в некоторых окрестностях.

Проблема состоит в следующем. Если эти возмущения единичны, то можно считать, что они не меняют степени однородности. Если же они начинают приобретать массовый характер, то система выходит из состояния равновесия. По-видимому, можно путем перераспределения слов по окрестностям (т. е. укрупнения или, наоборот, уменьшения их) некоторым нетривиальным способом повысить степень однородности языка.

Решение этой проблемы позволит построить динамическую модель, основанную на некоторых кибернетических соображениях и позволяющую показать (разумеется, в упрощенном, моделирующем виде), как происходит переход от одного состояния к другому при развитии системы языка.

Так, можно представить себе определенный этап развития древних индоевропейских диалектов, при котором система имени отличалась высокой степенью однородности, поскольку можно считать, что парадигмы единственного и множественного числа были почти не связанными между собой. Установление параллелизма единственного и множественного числа в ряде падежей приводит к перераспределению окрестностей. Возникает неоднородность, например связанная с отсутствием самостоятельной формы родительного падежа единственного числа. Но уже на почве отдельных индоевропейских языков достигается новое состояние устойчивости за счет введения — разными способами в разных языках — недостающих форм и тем самым заполнения пустых клеток.

Ясно, что подобная модель может быть построена лишь при наличии ясной картины синхронных срезов для разных периодов развития языка и построенных на основе знания соответствующих фактов обобщенных формальных моделей (в частности, должен быть выяснен тип неоднородности, степень ее устранимости, характер подсемейств и т. п.). Пока соответствующих данных еще не накоплено, но, тем не менее, поставить вопрос о конструировании динамических моделей необходимо.

§ 31. Замечания относительно морфемного анализа

В дальнейшем мы иногда будем пользоваться понятием «морфемы». Между тем, как было указано в § 17, формальное определение морфемы достаточно трудно. Мы будем исходить из содержательного определения, заключенного в следующих словах: «Это — наиболее элементарная неделимая далее значимая единица языка»¹³. Недостаток этого определения

¹³ П. С. Кузнецов. Значение грамматики для сравнительно-исторического языкознания. «Вопросы грамматического строя». М., 1955, стр. 114.

в том, что из него не следует, какой процедурой можно членить речевой поток на морфемы.

Поэтому для того формального аналога морфемы, который предстоит построить, необходимо определить некоторую процедуру выделения морфем. Ясно, что такой формальный аналог морфемы должен быть построен без непосредственного обращения к смыслу слов. Косвенное обращение к смыслу допускается в том классе моделей, который мы рассматриваем путем задания окрестностей. Мы видели также, что окрестности могут переопределяться на основании определенных внутрисистемных требований.

Теперь, исходя из окрестностей и семейств, довольно легко получить определение основы и флексии (вернее, их простейших аналогов, которые целесообразно разбирать в модели).

Возьмем некоторое слово x . Теперь мы уже не будем считать слово первичным понятием, т. е. чем-то неделимым, а представим его себе как некоторую последовательность фонем или, для простоты, букв: $l_1, l_2, \dots, l_{n-1}, l_n$.

Пусть наше слово x входит в $\Gamma_1 \cap S_1$.

Мы будем считать некоторый отрезок (возможно нулевой) флексией d_1 и оставшуюся часть основой f_1 , т. е. считать слово x имеющим вид $f_1 d_1$ при условии, что:

- 1) для всякого слова x' , входящего в $\Gamma(x)$ и имеющего вид $f_1 d_1$, найдется слово y , имеющее вид $f_1 d_1$ и входящее в $S(x')$;
- 2) для всякого слова y' , входящего в $S(x)$ и имеющего вид $f_1 d_1$, найдется слово x' , имеющее вид $f_1 d_1$ и входящее в $\Gamma(x')$.

Примечания:

1) Легко заметить, что здесь с некоторыми незначительными изменениями повторено определение, использованное для интерпретации 2 окрестности из § 24. Здесь, однако, предполагается, что окрестность взята в самой общей, третьей, интерпретации.

2) В приведенном здесь определении порядок следования частей f и d относительно друг друга можно не фиксировать, а рассматривать его как схему трех разных определений, а именно словоизменительного префикса (если d предшествует f), инфикса (если d вклинивается в f) и окончания. Мы для простоты будем иметь в виду лишь последний случай, особенно важный для большинства славянских языков.

Заметим теперь, что в нашем определении вовсе не требуется, чтобы все слова, входящие в данную окрестность, выделяли одну основу, а все слова, входящие в данное семейство, выделяли одно окончание. Это соответствует реальной ситуации, когда, например, *человек* и *люди* включаются в одну парадигму, а *время* и *поле* попадают в одно семейство. Подобные случаи, однако, достаточно редки. Введем поэтому еще следующие определения.

Мы будем называть окрестность правильной, если во всех словах этой окрестности выделяется одна и та же основа.

Мы будем называть семейство правильным, если во всех словах этого семейства выделяется одна и та же флексия.

Примечание. Правильные окрестности составляют подавляющее большинство окрестностей в конкретных языках, в то время как правильные семейства встречаются во флективных языках не очень часто. В агглютинативных языках правильных семейств большинство, но здесь, как мы видели в § 28, встречаются случаи непростоты, ведущие к неправильным семействам.

Поскольку мы не прибегали к значению, то могут спросить, почему мы не включаем в анализ такие слова, как *водка* и *водянка* (нарушается состояние дополнительного распределения), с одной стороны, и *водит*, *водворять*, *водவில்*, с другой. Дело, однако, в том, что эти слова, как легко видеть, не образуют столь стройных и богатых рядов со словами *водянистый* и другими рассмотренными нами, как в I и II рядах слов.

Последний пример показывает, однако, что, несмотря на правильность исходной идеи, необходимо преодолевать значительные трудности при анализе каждого случая.

Мы ограничиваемся этими замечаниями и не даем окончательного определения морфемы потому, что полная формализация морфемного анализа связана с построением развернутого статистического аппарата, а это выходит за рамки данной работы¹⁵.

В заключение выскажем некоторые соображения по поводу возможности морфемного членения слова на основании чисто статистических соображений без всякого обращения к понятию окрестности, которое, по-видимому, может быть получено на основании морфемного анализа.

Пусть мы получили разбиение на семейства, например:

S ₁	S ₂	S ₃
<i>серпом</i>	<i>серп</i>	<i>спорить</i>
<i>спором</i>	<i>спор</i>	<i>рулить</i>
<i>рулем</i>	<i>руль</i>	<i>пахать</i>
<i>пахарем</i>	<i>дом</i>	и т. д.
<i>ромом</i>	<i>ром</i>	
<i>веществом</i>	<i>ком</i>	
<i>болотом</i>	и т. д.	
и т. д.		

Будем теперь делить каждое слово на две части, левую и правую, например:

<i>нуль/серпом</i>	<i>серп/ом</i>
<i>с/ерпом</i>	<i>серпо/м</i>
<i>се/рпом</i>	<i>серпом/нуль</i>
<i>сер/пом</i>	

Сразу же остановимся на таких разделениях, при которых в правой части выделяется в достаточно большом числе случаев общий элемент для разных слов семейства.

Такие разделения мы назовем морфологическими. Например, для слова *серпом* из S₁ морфологическими будут разделения *серп/ом* и *серпо/м*. Заметим, что для нас оба эти разделения равноправны, так как оба при надлежащем выборе N характерны для достаточно большого числа случаев.

Для слова *серп* единственным морфологическим разделением будет разделение *серп/нуль*, но для слова *руль* возможны два: *руль/нуль* и *рул/ь*¹⁶.

¹⁵ Очень интересна в этом отношении модель, предложенная Н. Д. Андреевым и удачно сочетающая статистические методы с аппаратом теоретико-множественной концепции языка, см.: Н. Д. Андреев. Моделирование языка на базе его статистической и теоретико-множественной структуры. «Тезисы совещания по математической лингвистике». Л., 1959, стр. 15—22.

¹⁶ Ясно, что на фонетическом уровне морфологическое разделение проходит более естественно, но нас здесь интересует лишь общий принцип.

Возьмем теперь в каком-нибудь семействе (например, в S_1) то морфологическое разделение, которое проходит ближе к центру слова (например, *сери(ом)*). Если при выбранном морфологическом разделении для слов данного семейства найдется такое морфологическое разделение для некоторого другого семейства (например, S_2), что для достаточно большого числа левых частей (в нашем случае — для *сери-*, *спор-*, *рул-*, *ром-*) найдется совпадающая с ней левая часть (в нашем случае — в S_2), и наоборот: для достаточно большого числа левых частей во втором семействе (например, в S_2) найдется совпадающая левая часть в исходном семействе, то мы будем говорить, что эти семейства с в я з а н ы между собой. Если два слова с общей левой частью принадлежат к двум связанным семействам, то мы будем считать, что эти слова принадлежат одной окрестности. Ясно, что таким образом мы можем объединить в одну окрестность *серном* и *сери*, *рулем* и *руль*, *ромом* и *ром*, *спором* и *спор*. Ясно также, что *спорить* и *спор* не попадут в одну окрестность, поскольку семейства S_2 и S_3 не являются связанными. В самом деле, мы потребовали, чтобы для достаточно большого числа слов из S_2 нашлась при данном разделении совпадающая часть из S_3 : для *сери* не найдутся слова *сериить* или *сериать*; то же верно относительно слов *дом*, *ром*, *ком*.

Данный метод нуждается, конечно, в уточнении, например для «беглых гласных» в словах типа *осел* (род. п. *осла*), но в целом он вполне обеспечивает объединение слов в окрестности. Его можно усовершенствовать и в другом направлении, а именно для случая, когда изменяется не правая, а левая часть слова (для выделения префиксов).

Основная трудность состоит, по-видимому, в нахождении числа N достаточно большого, чтобы получить нужные морфологические разделения. Но, по-видимому, оно может быть найдено экспериментальным путем.

Важно, однако, что понятие окрестности в принципе может быть выведено, и достаточно задания всего двух исходных понятий: «слово» и «отмеченная фраза».

§ 32. Проблема грамматической омонимии

В дальнейшем (§ 36) будет сделана попытка описать некоторую модель распознавания, т. е. перехода от единиц текста к единицам системы. Для ранее рассмотренных моделей, где явно или неявно предполагалась регулярность языка, т. е. отсутствие омонимов, построение такой модели было бы тривиальностью. Поэтому в дальнейшей части этой главы предметом нашего рассмотрения будет нерегулярный язык. Мы исследуем модель, в которой омонимия определяется через свойство эквивалентности, а затем обсудим вопросы, связанные с построением модели распознавания.

Переходим к модели, предложенной Р. Л. Добрушиным.

Модель Добрушина возникла из рассмотрения гипотезы о том, что «семейство» соответствует «надежу» (см. § 23).

Сразу же оказалось, что это, вообще говоря, не так. Например, оказалось, что в русском языке выделяются три разных семейства:

S_1 — *окно*, *солнце*, *тело*, *туловище*, *око*...

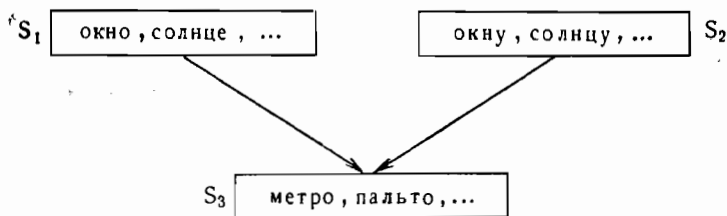
S_2 — *окну*, *солнцу*, *телу*, *туловищу*, *оку*...

S_3 — *метро*, *пальто*...

Между тем нам хотелось бы, чтобы слова из S_3 были как-то связаны со словами из S_1 и со словами из S_2 .

В связи с этим была предложена следующая система определений¹⁷. Мы будем говорить, что семейство S_i подчиняет семейство S_j ($S_i \rightarrow S_j$), если для любого слова x , входящего в S_i , и для любого слова y , входящего в S_j , из отмеченности фразы $A_1 x A_2$ следует отмеченность фразы $A_1 y A_2$ (где A_1 и A_2 — произвольные группы слов). Семейство S_k называется *начальным*, если не существует семейства S_e , такого, что $S_e \rightarrow S_k$ и $S_e \neq S_k$.

С каждым семейством сопоставляется совокупность всех слов, входящих в него и во все подчиненные ему семейства. Эту совокупность мы будем называть *элементарным грамматическим рядом*. (Отметим, что Р. Л. Добрушин называет ее элементарной грамматической категорией. Однако термин «категория» мы уже использовали, причем в значении, более принятом в лингвистике — см. § 25—26.) Изобразим подчинение при помощи следующей схемы:



Мы получили два элементарных ряда:

- 1) *окно, солнце, метро, пальто...*
- 2) *окну, солнцу, метро, пальто...*

Данное понятие позволяет решить ряд вопросов, связанных с формальным различием морфологических омонимов.

Например, слово *печь* входит как в категории имени, так и категории глагола. Система определений, предложенных Р. Л. Добрушиным, дает возможность формального определения принадлежности слова *печь* двум видам категорий. В самом деле, мы имеем следующие семейства:

- S_4 — *голь, сталь, мразь, речь...*
- S_5 — *стеречь, доить, пускать...*
- S_6 — *печь, течь...*

¹⁷ Р. Л. Добрушин. Элементарная грамматическая категория «Бюллетень объединения по проблемам машинного перевода», № 5, М., 1957, стр. 19—20.

Легко видеть, что здесь получаются два элементарных ряда:

3) *голь, сталь, мразь, печь, течь...*

4) *стеречь, доить, пускать, печь, течь...*

Естественно назвать класс слов морфологически омонимичным, если он входит в два (или более) элементарных ряда.

Таким образом, мы получили, не прибегая к значению, формальное определение морфологического омонима.

Этот результат чрезвычайно важен именно потому, что здесь не использовалось ни понятие окрестности, и полученное с его помощью понятие подсемейства, ни понятие категории.

В терминах принадлежности к категориям, разумеется, понятие омонима определить очень легко. Но в нашей системе определение омонима как слова, соответствующего дизъюнкции элементарных категорий, — определение, вообще говоря, вполне разумное — привело бы к логическому кругу, поскольку задание категорий исследовалось нами в предположении о регулярности языка, т. е. об отсутствии омонимов.

Теперь же, поскольку определение омонима построено нами независимо от задания категорий, мы можем считать, что сначала действует модель, снабжающая омонимами типа *пальто* или *печь* разными индексами в зависимости от принадлежности к разным элементарным рядам. Затем мы поставим вопрос о распространении модели § 29 на нерегулярный язык (см. § 34). Но прежде мы сделаем одно небольшое отступление.

§ 33. О связи между синкретизмом и нейтрализацией

В свое время Ельмслев высказал замечательную гипотезу о том, что явления нейтрализации в фонологии (в «плане выражения») и синкретизма в грамматике (в «плане содержания») суть два проявления одного и того же явления¹⁸.

В последнее время предпринят ряд попыток обобщения понятия нейтрализации. Наиболее интересны в этом отношении, по-видимому, идеи, высказанные чешским лингвистом Б. Трнка¹⁹. Наряду со случаями «синтагматической нейтрализации» (так Трнка называет те явления, которые мы разбирали для фонологии в главе II и для грамматики в главе III) предлагается рассмотреть случаи «парадигматической нейтрализации»,

¹⁸ См.: Л. Ельмслев. Прологомены к теории языка. «Новое в лингвистике», вып. I. М., 1960, стр. 344—345.

¹⁹ В. Трнка. On some problems of neutralization. «Omăgiu lui Jorgu Jordan cu prilejul împlinirii a 70 de ani». Bukurești, 1958, стр. 861—866.

т. е. случаи, когда совпадают члены двух парадигм, соответствующие двум разным элементарным категориям. Например, в русском языке в семейство, рассмотренное в § 29: *коровам, лампам, столам, любям, окнам, телам* и т. п., входят слова, соответствующие категориям мужского, женского и среднего рода, одушевленности и неодушевленности. Вслед за Трнкой можно сказать, что в данной парадигматической позиции нейтрализуется противопоставление между соответствующими категориями.

Трнка предлагает терминологически отграничить эти случаи от случаев, где неразличение имеет место не для всех парадигм. Так, например, слово *ночи* соответствует родительному, или дательному, или предложному падежу, единственному или множественному числу и т. п. Однако в парадигме слова *нога* соответствующие категории различаются. Этот случай Трнка и предлагает назвать случаем омонимии в отличие от парадигматической нейтрализации, рассмотренной выше.

Наши модели показывают, что здесь мы действительно имеем дело с двумя различными случаями.

Случай омонимии в смысле Трнки соответствует тому случаю, когда одно семейство, например то, в которое входит слово *ночи*, принадлежит двум и более элементарным рядам (или же рядам, представляющим некоторое обобщение этого понятия — см. § 34). Как мы видели, для разграничения омонимии в этом случае достаточно построить простейшую модель, не использующую иных исходных понятий, кроме «слова» и «отмеченной фразы».

Случай парадигматической нейтрализации может быть отражен лишь в более сложной модели, требующей введения понятий «окрестность» и «категория». Тогда, как мы видели в § 30, семейство *коровам, лампам, столам, окнам* и т. п. можно расчленил на ряд подсемейств, каждое из которых соответствует одной элементарной категории.

Посмотрим теперь, в какой степени может быть распространена на случай языка с омонимией модель, ставящая подсемейства в соответствие с элементарными категориями.

§ 34. Соответствие между грамматическими категориями и элементарными рядами

Принцип связи между окрестностями и категориями, выдвинутый в § 25, зафиксировал, что в случае, когда (хотя бы в простом языке) два слова входят в одну окрестность, они должны относиться к разным элементарным категориям.

Если теперь существует хотя бы одна такая окрестность, содержащая слова x_1 и x_2 , то любое слово y , входящее в один элементарный ряд с x_1 и в один элементарный ряд с x_2 , целе-

сообразно считать относящимся (хотя бы для одной пары категорий) одновременно к двум разным категориям.

На основании применения этого принципа к категории падежа можно прийти к выводу, что, кроме выделяемых шести падежей в русском языке, необходимо выделить еще и следующие:

- 1) местный падеж: *в лесу, в году*;
- 2) количественно-разделительный падеж: *выпить чаю, дать воды* и т. п.

Это, впрочем, вполне соответствует мнению целого ряда грамматистов (например, А. А. Шахматова, В. А. Богородицкого, А. М. Пешковского).

Основная трудность связана здесь со следующим обстоятельством. Наш принцип выделения новых категорий чисто логический: требуется существование «хотя бы одного слова, такого, что...». Когда мы переходим к интерпретации модели, то, по-видимому, логический принцип должен быть заменен статистическим²⁰.

Рассмотрим такой пример. В чешском языке супин не отличается от инфинитива во всех глаголах, кроме глагола *spáti*: инфинитив *spát* (с долгим *a*), а супин *spat*. Значит ли это, что в любом другом инфинитиве чешского языка, например *číst*, мы должны видеть омоним, принадлежащий двум категориям? Любой лингвист ответит на этот вопрос отрицательно.

По-видимому, непризнание приведенных падежей объясняется статистическими, вернее интуитивно-статистическими, соображениями.

У этого вопроса есть и другая сторона. В § 26 мы указывали, какое значение имеет при задании категорий условие категориальной однородности. Поэтому следует стремиться к тому, чтобы задание новых категорий не приводило к увеличению категориальной неоднородности. Это требование можно, в частности, выполнить, если ограничиться некоторым подязыком.

Так, если выделить для слов неодушевленно-мужского рода в русском языке местный и качественно-разделительный падежи, то для категориальной однородности было бы необходимо, чтобы соответствующие падежи задавались и для остальных родов русского языка. Между тем, в большинстве случаев для этого не существует формальных показателей.

Поэтому мы предпочитаем в этом пункте остаться на точке зрения традиционной грамматики, выделяющей всего шесть падежей в русском языке. Оговариваюсь, что мы выбрали для применения своей модели подязык, в котором такие слова,

²⁰ Этот вопрос подробно разбирается в работе: С. Стати. Омонимия в морфологической системе. «*Revue de linguistique*», t. V, 1960, № 2, стр. 257—263.

как *лес*, отсутствуют. Тем самым остается открытым путь для такого моделирования, при котором для каждого подязыка строится своя частная падежная система (имеется в виду понятие, введенное Ельмслевом)²¹. При таком подходе был бы бессмысленным вопрос, сколько, например, падежей вообще в русском языке, поскольку ответ на подобный вопрос считался бы возможным только для специально отобранных подязыков.

После этих предварительных замечаний мы можем сформулировать требование относительно задания категорий, аналогичное требованию регулярности (§ 29). Назовем те подсемейства, которые входят в начальные семейства (§ 32), начальными подсемействами. Мы будем называть язык (с заданной системой категорий) почти регулярным, если можно так отобразить множество слов на множество категорий, чтобы полный прообраз каждой элементарной категории совпал с некоторым начальным подсемейством. Для такого языка соответствие между категориями и классами слов, вошедших в начальные подсемейства, устанавливается методом, описанным в § 29. Разбив затем все слова на морфологические подклассы, можно отнести слова, не вошедшие в начальные подсемейства, к тем парадигматическим категориям, к которым отнесены слова из того же морфологического подкласса, вошедшие в начальные подсемейства, и к тем непарадигматическим категориям, к которым отнесены слова из начального семейства. Остается, правда, неясным, получим ли мы таким образом задание категорий, которое соответствует нашему представлению об определенном языке.

Так, сразу же видно, что этой процедурой некоторую новую категорию падежа мы выделим лишь в том случае, если в некоторой окрестности Γ существует слово x , входящее только в этот падеж. Или то же на обычном лингвистическом языке: необходимо, чтобы существовала хотя бы одна парадигма, в которой данная падежная форма не совпадала бы с какой-нибудь другой.

Необходимость этого условия иллюстрируется следующим примером:

Пусть в языке есть четыре падежа и две парадигмы склонения (a, b, c, d, e, f словоформы).

	I	II
Им. п.	a	d
Род п.	b	f
Дат. п.	c	e
Вин. п.	a	e

²¹ L. Hjelmslev. La catégorie de cas. «Acta jutlandica», VII. Aarhus, 1935, стр. 84.

В этом случае мы (при некоторых условиях, о которых речь будет идти ниже) могли бы получить элементарный ряд (a, d).

Но мы никогда не получим элементарного ряда (a, e), так как нет такого семейства, которое (при фиксированном роде и числе) можно было бы поставить в соответствие с винительным падежом²².

Подобные случаи омонимии достаточно редки, но все же не исключены. Например, в современном армянском языке винительный падеж для названий лиц совпадает с дательным, а для неодушевленных предметов — с именительным. В древнеанглийском языке в основах на *-o* (например, типа *пато* 'имя') в единственном числе винительный совпадает с дательным, для остальных слов в единственном числе, а также для всех слов во множественном числе винительный совпадает с именительным.

Впрочем, за примерами не нужно идти так далеко. Подобная же картина наблюдается и для некоторых фрагментов русского языка (а также и других славянских языков).

Рассмотрим следующие семейства русского языка:

S_1 — пахарь, вратарь, брат, отец, профессор...

S_2 — серп, молот, свинец, карандаш, чайник...

S_3 — серпа, молота, свинца, карандаша, чайника...

S_4 — пахаря, вратаря, брата, отца, профессора...

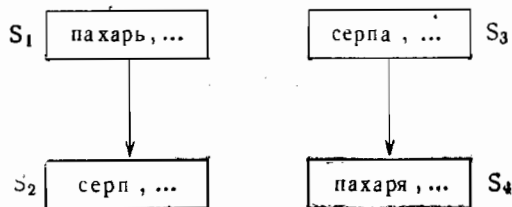
Мы хотели бы, чтобы эти слова распределялись по трем элементарным рядам:

I — им. п. пахарь, вратарь, серп, молот...

II — род. п. серпа, молота, пахаря, вратаря...

III — вин. п. пахаря, вратаря, серп, молот...

Однако применением определений § 28 мы получим следующую схему подчинения:



²² Это становится ясным из следующих рассуждений. Такое распределение форм в окрестностях в терминах нашей абстрактной системы можно выразить так: $S(b) = S(f)$; $S(d) \rightarrow S(a)$ и $S(c) \rightarrow S(e)$. В этих условиях, как легко проверить, не может иметь места ни $S(a) \rightarrow S(e)$, ни $S(e) \rightarrow S(a)$. Пусть, например, $S(a) \rightarrow S(e)$. Тогда (по транзитивности) $S(d) \rightarrow S(e)$, а это может быть лишь тогда, когда всюду вместо d может стоять e . Мы же предположили, что d и e — разные словоформы.

т. е. мы можем получить только два элементарных ряда вместо ожидаемых трех, винительный исчез. Пропала всякая связь между семейством S_2 (*серп, молот...*) и семейством S_4 (*пахаря...*). В самом деле, не всегда можно заменить *серп* на *пахаря* (например, нельзя этого сделать во фразе: «*серп* лежит на полу»), и не всегда можно заменить *пахаря* на *серп* (нельзя, например, во фразе «труд *пахаря* — самый почетный»).

Этот пример помогает уяснить также то обстоятельство, что выдвинутое нами и, как мы видели, необходимое условие существования хотя бы одной парадигмы, в которой данная форма неомонимична какой-нибудь другой, не является достаточным. В самом деле, в разобранном фрагменте русского языка совпадают в одной парадигме винительный и именительный падежи, а в другой родительный и винительный. В русском языке есть, однако, парадигма склонения (*корова, коровы, корове, корову* и т. п.), где соответствующие падежи не омонимичны. Дело, однако, в том, что элементарные категории для *корову* и *пахаря* не совпадают (различны категории рода) и поэтому между соответствующими семействами нет связи. Надо, очевидно, потребовать наличия для каждой элементарной категории такой парадигмы, в которой некоторое слово соответствовало бы только ей.

Можно идти и другим путем. Если устранить все согласуемые с существительными слова, то слово *корову* станет начальным семейством для слов *пахаря* и *серп*. Больше того. Можно, оказывается, построить модель, в которой, во-первых, ограничение множества отмеченных фраз производится формально и которая, во-вторых, позволяет разграничить, например, падежи в случае, когда выдвинутое нами условие вообще не выполняется.

§ 35. Расширение модели элементарного ряда

Введем ряд новых определений. Назовем S_1 -фразой отмеченную фразу, в которой замена хотя бы одного слова на слово из семейства S_1 составляет фразу отмеченной. В частности, S_1 -фразами являются все фразы, содержащие хотя бы один элемент из семейства S_1 .

Очевидно также следующее: Если $S_1 \rightarrow S_j$, то все S_1 -фразы одновременно являются S_j -фразами.

Рассмотрим теперь множество отмеченных фраз, такое, что в каждой отмеченной фразе встречается не больше одного слова из данного элементарного ряда (ясно, что такое ограничение вполне оправдано, например, при рассмотрении проблемы падежа).

Язык, в любой отмеченной фразе которого встречается не больше одного слова из данного элементарного ряда, назовем элементарным языком.

Теорема 4.7. В элементарном языке из условия $S_1 \rightarrow S_j$ следует, что хотя бы одна S_j -фраза не является S_1 -фразой.

Пусть теперь $S_j \rightarrow S_i$. Объявим все S_i -фразы неотмеченными, т. е. временно удалим их из множества отмеченных фраз. Как мы видели, в этом случае останется хотя бы одна S_j -фраза. Можно показать, что на оставшемся множестве фраз два слова, принадлежавшие ранее одному семейству, останутся в одном семействе.

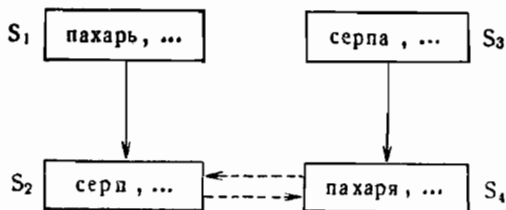
При этом имеются в принципе две возможности:

1) S_j окажется подчиненным какому-то семейству S_k . Тогда мы будем говорить, что S_j условно подчинено S_k .

2) S_j не подчинено никакому семейству. Этот случай может быть описан при помощи видоизмененной системы определений, предложенной Добрушиным позднее²³. Здесь мы исключаем эти случаи из рассмотрения.

Совокупность слов, входящих в какое-то семейство и во все подчиненные ему или во все условно подчиненные ему, мы будем называть грамматическим рядом (таким образом, грамматический ряд — понятие более общее, чем элементарный грамматический ряд).

Возьмем теперь семейство S_2 (*серп, молот...*), подчиненное семейству S_1 (*пахарь, вратарь*). Исключим все S_1 -фразы, например: *Пахарь* пахнет; *Пахарь* выходит в поле; *Пахарь* любит селедку... Теперь семейство S_2 (*серп, молот...*) становится условно подчиняющим семейство S_4 (*пахаря...*). В самом деле, во всех оставшихся отмеченных фразах, например: *Надо любить серп*; *Там найдут серп*; *Куда положат серп?* и других можно заменить *серп* на *пахаря*. С другой стороны, после удаления всех S_3 -фраз типа «ручка *молота*», «голова *серпа*» семейство S_2 (*серп*) становится условно подчиненным семейству S_4 (*пахаря*). Получаем следующую схему (линиями указано подчинение, пунктиром условное подчинение):



Получаем следующие грамматические ряды:

- а) *пахарь, серп, молот...*
- б) *серпа, молота, пахаря...*
- в) *серп, молот, пахаря...*
- г) *пахаря, серп, молот...*

Поскольку два последних ряда совпадают, то мы получили три ряда, соответствующие именительному, родительному и винительному падежам.

§ 36. Замечания о построении распознающей модели

Пусть теперь язык представлен как совокупность категорий и окрестностей, а речь как некоторое множество отмеченных фраз. Наша задача состоит в том, чтобы каждое слово поставить в соответствие с некоторой элементарной категорией.

²³ Р. Л. Добрушин. Математические методы в лингвистике. Приложение, «Математическое просвещение», вып. 6. М., 1961, стр. 52—59.

В отличие от методов, применявшихся ранее, для анализа некоторой фразы мы обращаемся не ко всему множеству фраз, а используем сведения, добытые в ранее рассмотренных моделях. Мы видели (§ 31), что принадлежность к данному семейству или подсемейству определяется окончанием, причем если для простоты ограничиться правильными семействами, то некоторое окончание и соответствует информации о семействе.

Таким образом наша задача состоит в том, чтобы:

- 1) расчленить каждое слово фразы на основу и окончание;
- 2) сопоставить окончание с некоторым семейством;
- 3) сопоставить семейство с некоторой элементарной категорией (второй и третий этап легко могут быть объединены).

При этом предполагается, что у нас есть таблица окончаний, список семейств и элементарных категорий.

При несколько другой терминологии математическая сторона возникающей здесь задачи была подробно исследована С. Я. Фиталовым²⁴.

В этой — очень интересной именно с точки зрения моделирования — работе показано, что каждая из этих операций может быть представлена некоторой функцией.

Наиболее интересны замечания о характере функции, ставящей в соответствие с окончанием некоторую элементарную категорию (функция Φ).

Как мы знаем, семейству, вообще говоря, соответствует некоторая дизъюнкция элементарных категорий, поэтому некоторое окончание α делает необходимым определение ряда функций $\Phi_1(\alpha)$, $\Phi_2(\alpha)$ и т. п. Естественно поэтому определить новую функцию $\varphi(\alpha)$, имеющую вид:

$$\varphi(\alpha) = \Phi_1(\alpha) \cup \Phi_2(\alpha) \cup \dots \cup \Phi_n(\alpha).$$

Далее в работе исследуются требования, которые следует наложить на такого рода функции, но это исследование носит уже чисто математический характер. Нас интересует здесь лишь следующее обстоятельство. Построение распознающей модели существенно зависит от того, насколько распространена в данном языке омонимия и парадигматическая нейтрализация. Относящиеся сюда исследования представляют и самостоятельный лингвистический интерес и могут, по-видимому, быть использованы для типологических исследований.

Для чешского языка подобное исследование было произведено П. Сгаллом²⁵, предложившим и количественные методы оценки явлений омонимии.

²⁴ С. Я. Ф и т а л о в. О построении формальной морфологии в связи с машинным переводом. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста», вып. 2. М., 1961.

²⁵ P. S g a l l. Soustava pádových koncovek v češtině. «Acta Universitatis Carolinae. Philologica». Praha, 1960, 2, стр. 65—84.

Остановимся на этой работе. Автор по понятным причинам отдельно выделяет степени сравнения (где одна морфема соответствует не элементарной грамматической категории, как это обычно бывает во флективных языках, а отдельной категории, как это имеет место в языках агглютинативных). Если отвлечься от них, то всего в чешском языке теоретически возможны 452 элементарных категории как разные комбинации 14 категорий. Если бы между элементарными категориями и окончаниями имелось бы взаимнооднозначное соответствие, то в чешском языке было бы 452 разных окончания. На самом деле это далеко не так. Парадигматические нейтрализации исключают 394 случая. Остается таким образом 58 разных морфем. Практически разных окончаний (звуковых оболочек окончаний) 34, но если считать два совпадающих окончания, соответствующие разным элементарным категориям, разными типоформами²⁶, то число таких типоформ 116.

Предлагается рассмотреть следующие характеристики языка:

1) отношение числа типоформ к числу морфем, т. е. для чешского языка $\frac{116}{58} = 2$; этот показатель оценивает степень синонимичности окончаний;

2) отношение числа типоформ к числу разных окончаний, т. е. для чешского языка $\frac{116}{34} = 3,41$; этот показатель оценивает степень омонимичности окончаний.

Пока еще отсутствуют данные для сравнения соответствующих показателей других языков, но надо надеяться, что эта методика окажется полезной.

²⁶ Термин «типоформа» заимствован из работы А. И. Смирницкого «Лексическое и грамматическое в слове». — «Вопросы грамматического строя», М., 1955, стр. 36.

СИНТАГМАТИЧЕСКИЕ МОДЕЛИ
В ГРАММАТИКЕ

§ 37. Модель, перенесенная из фонологии

Простейшим способом построения синтагматической модели на синтаксическом уровне является перенесение в синтаксис модели, которую мы обсуждали в § 15.

Возьмем все фразы языка («отмеченные фразы»). Снабдим каждое слово индексом, соответственно с местом, которое оно занимает в некоторой фиксированной фразе. Слова с разными индексами места будем считать разными элементами. Возьмем некоторое слово $a_{\alpha i}$, где α — номер слова в словаре, а i — номер места в некоторой фразе. Составим все пары $a_{\alpha i} a_{\beta i+1}$, которые действительно встречаются. Мы получим граф, характеризующий некоторый язык. Любая совокупность фраз дает возможность построить подграф этого графа. Например, пусть таны фразы:

Человек₁ человеку₂ волк₃,
Человек₁ ходит₂ прямо₃,
Человека₁ убивает₂ волк₃.

Получим следующий подграф:



Подобная модель позволяет решать те же вопросы, что и соответствующая фонологическая модель (проблемы типологического сравнения языков по некоторым заданным характеристикам, возможности восстановления промежуточных слов и т. п.).

Однако эта модель не может адекватно описать явления синтаксиса, поскольку в ней устанавливаются отношения лишь между рядом стоящими словами, ср., например:

в → *очень* → *большой* → *дом*,

где пара (*в*, *очень*) отражает совершенно случайную связь.

Неадекватность модели станет еще более очевидной в следующем параграфе, где мы рассмотрим соответствующую ей модель порождения (язык с конечным числом состояний).

Тем не менее эта модель вполне может быть использована для описания такого множества фраз, где каждое соположение отражает некоторую внутреннюю связь.

§ 38. Язык с конечным числом состояний

Построим теперь синтетическую модель, обратную по отношению к модели, данной в § 37.

Пусть имеется некоторое механическое устройство, способное принимать конечное число состояний:

$$C_0, C_1, \dots, C_n.$$

Работа этого устройства состоит в том, что оно переходит из начального состояния C_0 в какие-то состояния C_i , а затем устройство проходит некоторую последовательность состояний:

$$C_{i_1}, C_{i_2}, \dots, C_{i_k}.$$

Если устройство переходит из какого-нибудь состояния C_j в состояние C_k , то эта пара состояний называется связанной и при этом переходе выдается слово a_{jk} . При переходе из некоторого состояния в начальное выдается знак конца λ (например, точка). Таким образом, каждая последовательность состояний, проходимых устройством, соответствует некоторому кортежу слов. Если эта последовательность состояний такова, что первое и последнее состояния суть C_0 , то соответствующий этой последовательности кортеж слов называется отмеченной фразой, а все множество таким образом порожденных отмеченных фраз называется языком с конечным числом состояний.

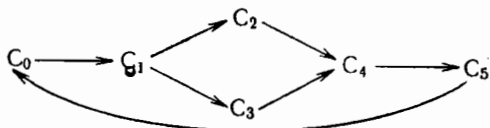
Если в какой-нибудь последовательности состояний, соответствующей отмеченной фразе, некоторое нена начальное состояние проходится дважды (т. е. если при порождении некоторой отмеченной фразы устройство возвращается в то же самое нена начальное состояние), то говорят, что устройство имеет цикл. Легко видеть, что устройство, имеющее хотя бы один цикл, может порождать отмеченные фразы сколь угодно большой длины. Таким образом, язык с конечным числом состояний может состоять из бесконечного числа фраз.

Необходимо отметить, что состояния, принимаемые устройством, и все возможные переходы задаются извне. Мы хотим так задать состояния и так связать переходы из одного состояния в другое со словами исследуемого языка, чтобы отмеченные фразы, порождаемые устройством, совпадали с фразами этого языка. Покажем теперь на некоторых фрагментах русского языка, каким образом можно было бы выбрать состояния и поставить переходы в соответствие со словами русского языка, чтобы представить эти фрагменты как язык с конечным числом состояний.

Так, например, отмеченные фразы:

Где качались тихо ели,
Где шумели звонко ели...
(Хлебников)

могут быть произведены следующим устройством, принимающим шесть состояний:



Связанной паре C_0C_1 соответствует слово *где*

»	»	C_1C_2	»	»	<i>качались</i>
»	»	C_1C_3	»	»	<i>шумели</i>
»	»	C_2C_4	»	»	<i>тихо</i>
»	»	C_3C_4	»	»	<i>звонко</i>
»	»	C_4C_5	»	»	<i>ели.</i>

Возьмем теперь начало стихотворения Брюсова «Сухие листья»:

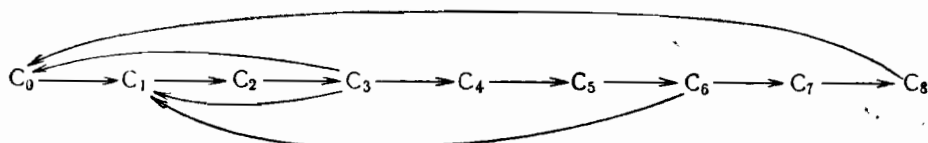
Сухие листья, сухие листья,
сухие листья, сухие листья
Под тусклым ветром кружат, шуршат.
Сухие листья, сухие листья
Под тусклым ветром сухие листья...

Эти две фразы (мы будем считать точку знаком конца) могут быть порождены следующим устройством, принимающим всего восемь состояний:

Связанной паре C_0C_1 соответствует пробел

»	»	C_1C_2	»	слово <i>сухие</i>
»	»	C_2C_3	»	» <i>листья</i>
»	»	C_3C_0	»	знак конца
»	»	C_3C_4	»	слово <i>под</i>

Связанной паре C_4C_5	соответствует слово	<i>тусклым</i>
»	» C_5C_6	» <i>ветром</i>
»	» C_6C_1	» <i>пустое слово</i>
»	» C_6C_7	» <i>кружат</i>
»	» C_7C_8	» <i>шуршат</i>
»	» C_8C_0	» <i>знак конца</i>
»	» C_3C_1	» <i>запятая</i>



Легко заметить следующее. Кроме указанных двух фраз, наше устройство произведет бесконечное множество других фраз, например, фраз вида:

сухие листья, сухие листья... сухие листья

n раз

Таким образом, мы видим, что язык с конечным числом состояний может, если допускаются циклы, содержать бесконечное число отмеченных фраз.

Схема языка с конечным числом состояний, предложенная Н. Хомским и изложенная выше, имеет с нашей точки зрения один существенный недостаток. А именно, мы установили, что каждой связанной паре (C_i, C_j) соответствует одно слово $a_{ij\alpha}$. Однако одному и тому же слову a_α могут, вообще говоря, соответствовать разные связанные пары. Это делает неудобным применение схемы к анализу конкретных языков. В самом деле, выше была описана схема порождения двух строк хлебниковского стихотворения. Но почему должна была быть выбрана именно эта схема? Ведь мыслима и следующая схема:



Можно, разумеется, потребовать, чтобы данное множество фраз описывалось путем задания минимального из возможных чисел состояний. Мы наложим, однако, гораздо более слабое ограничение на задание состояний. Это ограни-

чение ничего не говорит о способе порождения фраз типа описанных выше двух схем, но оно, как мы увидим, имеет определенный лингвистический смысл, соответствуя нашей грамматической интуиции. Обозначим каждое задаваемое состояние через $C_{\alpha\beta}$, где α — номер в словаре какого-нибудь слова, которое производится при переходе в это состояние, а β — номер в словаре какого-нибудь слова, которое производится при переходе из этого состояния в какое-либо другое. Потребуем теперь, чтобы при порождении двух произвольных фраз, содержащих на одинаковом расстоянии от начала (измеряемом количеством предшествующих слов) одну и ту же пару рядом стоящих слов $a_\alpha a_\beta$, проходило состояние $C_{\alpha\beta}$.

Назовем множество фраз, порожденное схемой с указанным ограничением, ограниченным языком с конечным числом состояний.

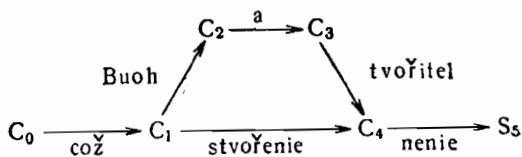
Наложенное нами ограничение действительно весьма слабое. Мы хотели бы, например, чтобы сочетание *хороший человек* однозначно определяло некоторое состояние $C_{\text{хороший, человек}}$. Но ведь в русском языке такое сочетание может стоять во фразе на любом месте, а мы потребовали лишь однозначности прохождения состояния лишь для сочетаний, стоящих на одинаковом расстоянии от начала. Возьмем строфу:

*Což Buoh a tvořitel nenie,
tot' všeko slove stvořenie,
čož stvořenie nenie,
tot' jest stvořitel!*

(из гуситских песен)

«Что не Бог и творец,
то некоего слова творение,
что не есть творение,
то есть творец!»

Выделенные курсивом строки хотелось бы породить следующей схемой:

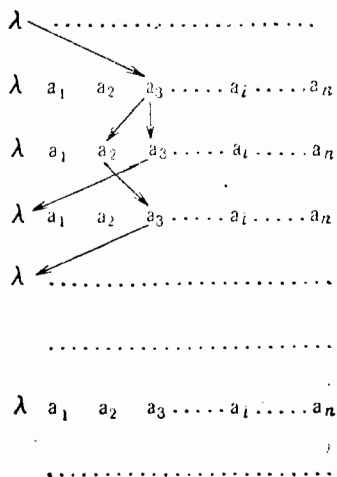


Однако при нашем ограничении подобная схема не обязательна, хотя с лингвистической и поэтической точки зрения нет особой разницы между этими строками и строками хлебниковского стихотворения.

Ограниченный язык с конечным числом состояний допускает интерпретацию, отличную от той, которая дается для всех

остальных языков с конечным числом состояний. Возьмем все слова языка (начиная со знака конца λ) и выпишем их в строчку. Повторим эту строчку сколь угодно большое число раз.

Каждая фраза на этом языке может быть представлена следующим образом. Слово, стоящее на i -м месте во фразе, берется из i -й строки. Стрелки соединяют слова, принадлежащие некоторой фразе. Например, показаны пути, которые проходятся для получения фраз $a_3a_2a_3$ и a_3a_3 .



Покажем теперь, что реальные языки, например все славянские языки, не являются ограниченными языками с конечным числом состояний.

Мы знаем, что в славянских языках порядок слов свободный. Поэтому существуют определенные связи между словами, стоящими на расстоянии друг от друга. Например:

Я, отрок, зажигаю свеч,
 Огонь кадильный берегу.
 Она без мысли и без речи
 На том смеется берегу

(А. Блок)

Здесь связаны между собой такие разделенные слова, как *она* и *смеется*, *на* и *берегу*.

Постараемся сформулировать теперь это свойство в достаточно абстрактных терминах.

Назовем некоторую отмеченную фразу $A = x_1 x_2 \dots x_n$ фразой с рамкой (i, j) , если в нее входят слова x_i и x_j ($j \geq i + 2$), для которых существуют слова y_i и y_j , такие, что фраза

$$x_1 \dots x_{i-1} y_i x_{i+1} \dots x_{j-1} x_j x_{j+1} \dots x_n$$

не отмечена, в то время как остается отмеченной фраза

$$x_1 \dots x_{i-1} y_i x_{i+1} \dots x_{j-1} y_j x_{j+1} \dots x_n,$$

иначе говоря, если замена x_i на y_i вызывает замену x_j на y_j .

Теорема 5.1. Язык, в который входит хотя бы одна фраза с рамкой (i, j) , не является ограниченным языком с конечным числом состояний.

Отсюда, между прочим, следует, что русский язык не является таким языком. Могут, правда, возразить, что приведенные случаи инверсии устранимы, поскольку достаточно описать формальным способом лишь фразы, не содержащие инверсию. Это возражение не кажется нам справедливым, поскольку свобода порядка слов есть существенный типологический признак славянских языков. Можно, однако, привести случаи и неустрашимых рамочных конструкций в русском (и любом другом славянском) языке, например *если..., то; не только..., но и* и т. п. В приложении показывается, что можно построить примеры языков, которые вообще не являются языками с конечным числом состояний (см. доказательство теоремы 5.11 на стр. 185).

§ 39. Конфигурационная модель Кулагиной

Недостатки рассмотренной аналитической модели, особенно ясно вскрывшиеся при рассмотрении соответствующей ей синтетической, объясняются в основном двумя причинами:

1) эта модель строится для конкретных фраз, а не для каких-нибудь структур, т. е. в ней не допускаются никакие мета-обозначения;

2) эта модель исследует лишь связи между рядом стоящими словами.

Теперь мы перейдем к рассмотрению модели, предложенной О. С. Кулагиной¹ для произвольных В-структур. Но в отличие от Кулагиной мы будем рассматривать не произвольные В-структуры, а некоторые специальные структуры. Дело в том, что очень трудно перейти от аналитической модели в произвольных В-структурах к соответствующей синтетической модели. В самом деле, зная, что некоторая В-структура отмечена, мы еще не можем сказать, отмечена ли некоторая конкретная фраза, отображающаяся в эту В-структуру. Отсюда очевидно, что синтагматическую модель целесообразно строить в совершенных структурах (см. § 22). Но существуют ли вообще совершенные В-структуры. На этот вопрос отвечает следующее утверждение.

Теорема 5,2: а) всякая отмеченная S-структура есть совер-

¹ О. С. Кулагина. Об одном способе определения грамматических понятий на базе теории множеств. «Проблемы кибернетики», вып. 1. М., 1958.

шенная В-структура; б) Если при некотором разбиении В всякая В-структура или не отмечена, или совершенна, то для любого элемента x верно, что В (x) есть некоторое подмножество семейства $S(x)$.

Эта теорема показывает, что синтаксис целесообразно строить именно в S-структурах, ибо семейство есть наиболее крупное объединение, при котором всякая структура совершенна.

В дальнейшем поэтому все описание будет вестись в терминах S-структур (хотя большинство определений и утверждений, делаемых относительно аналитических моделей, действительно для произвольных В-структур).

Для того чтобы рассматривать связи, существующие между словами, не обязательно расположенными контактно, в модели Кулагиной вводится понятие конфигурации определенного ранга.

Понятие конфигурации уточняет понятие конструкции и ее составляющих, как оно употребляется в дескриптивной лингвистике, а именно, конструкцией, как правило, называется группа слов, имеющая то же распределение, что и одно слово, а также любое высказывание².

При анализе фраз методом выделения конструкций и их составляющих большое значение приобретает понятие «ранга», введенное, по-видимому, еще Есперсеном³, который, правда, признавал всего три ранга. В структурной лингвистике понятие ранга было использовано, с одной стороны, де Гроотом⁴ и, с другой стороны, Харрисом⁵. При этом де Гроота интересовали отношения внутри некоторой группы слов, а Харриса — возможность путем введения соответствующей индексации расширить взаимозаменяемость слова и группы слов.

Все эти идеи отражены в модели, рассматриваемой ниже (хотя трудно предполагать, чтобы ее создатели были непосредственно знакомы с этим кругом идей структурной лингвистики).

Переходим теперь к основным понятиям.

Возьмем какую-то S-структуру. Назовем S-конфигурацией 1-го ранга (\tilde{S}_1) такую S-структуру, которая содержит не менее двух элементов и для которой существует элемент разбиения S (обозначим его S_{α_1}), такой, что структуры

$$S(A_1)\tilde{S}_1S(A_2) \text{ и } S(A_1)S_{\alpha_1}S(A_2)$$

² Правда, в книге Глиссона (Введение в дескриптивную лингвистику. Перевод с англ. М., 1959, стр. 190) конструкция определяется как «любая значащая группа слов», но дистрибутивный аналог этого понятия дать довольно просто.

³ См.: О. Есперсен. Философия грамматики. Перевод с англ., М., 1958, глава VII.

⁴ De Groot. Strukturelle Syntaxis. Den Haag, 1949; его же. Structural linguistics and syntactic laws. «Word», vol. 5, 1949, № 1.

⁵ Z. Harris. Methods in structural linguistics. Chicago, 1951, стр. 265—268.

являются одновременно отмеченными или неотмеченными при любых A_1 и A_2 . Элемент S_{α_1} называется результирующим для данной конфигурации. Заменяв все конфигурации 1-го ранга результирующими, мы получим S-структуру 1-го ранга (S_1).

В общем случае конфигурацией n-го ранга называется такая S-структура, состоящая не менее чем из двух элементов, для которой имеется элемент S_{α_n} , такой, что структуры (n-1)-го ранга

$$S(A_1)\tilde{S}_{(n)}S(A_2) \text{ и } S(A_1)S_{\alpha_n}S(A_2)$$

являются одновременно отмеченными или неотмеченными при любых A_1 и A_2 . Структурой n-го ранга называется структура, не содержащая конфигураций n-го ранга.

Определение конфигурации является индуктивным, или рекурсивным. Вместо того, чтобы определить отдельно S-конфигурации 2, 3, 25-го ранга мы сразу определили S-конфигурации n-го ранга (и S-структуры n-го ранга) через S-структуры и S-конфигурации (n-1)-го ранга. Эти в свою очередь определены через S-конфигурации и S-структуры (n-2)-го ранга и т. д. В конце концов мы дойдем до S-структур и S-конфигураций 1-го ранга, которые определены (базис индуктивного определения).

Разберем с этой точки зрения S-структуру:

$S(\text{маленькая})S(\text{девочка})S(\text{долго})S(\text{ласкала})S(\text{кошку})$.

Можно ли выделить в ней S-конфигурации 1-го ранга?

Попробуем заменить $S(\text{маленькая})S(\text{девочка})$ на $S(\text{девочка})$. Получим $S(\text{девочка})S(\text{долго})S(\text{ласкала})S(\text{кошку})$. Это тоже отмеченная S-структура. Можно ли, однако, считать S-структуру $S(\text{маленькая})S(\text{девочка})$ S-конфигурацией 1-го ранга? Определение требует, чтобы мы проверили возможность такой замены в любых окружениях.

Возьмем такое окружение:

$S(\text{весьма})S(\text{маленькая})S(\text{девочка})S(\text{стояла})$,

где $S(\text{весьма})$ — семейство всех наречий степени типа *очень*, *весьма* и т. п. $S(\text{стояла})$ — семейство непереходных глаголов типа *гуляла*, *играла*, *танцевала* и т. п.

Если мы в этой отмеченной S-структуре (отмечена она хотя бы потому, что есть фраза *Очень красивая дама стояла*), заменим $S(\text{маленькая})S(\text{девочка})$ на $S(\text{девочка})$, то получим:

$S(\text{весьма})S(\text{девочка})S(\text{стояла})$.

Эта S-структура не отмечена (если считать в русском языке фразы типа *Очень дама стояла* неотмеченными). Итак, $S(\text{маленькая})S(\text{девочка})$ не есть конфигурация 1-го ранга. Легко проверить, что $S(\text{весьма})S(\text{маленькая})$ — это конфигурация 1-го ранга. $S(\text{весьма})S(\text{маленькая})$ во всех окружениях может быть заменено на $S(\text{маленькая})$ и обратно, т. е. $S(\text{маленькая})$ есть

результатирующий конфигурации 1-го ранга (S_{x_1}). Если же мы будем рассматривать только S-структуры 1-го ранга, т. е. такие, в которых нет конфигураций 1-го ранга, то S(*маленькая*) S(*девочка*) всегда может быть заменено на S(*девочка*) и обратно.

Итак, S(*маленькая*)S(*девочка*) — это конфигурация 2-го ранга с результирующим $S_{x_2} = S(\text{девочка})$.

Будем анализировать нашу S-структуру дальше. S(*долго*) S(*ласкала*) — это конфигурация 2-го ранга с результирующим S(*ласкала*). Итак, структура 2-го ранга, полученная из исходной фразы, имеет вид:

$$S(\text{девочка})S(\text{ласкала})S(\text{кошку}).$$

Если рассматривать только S-структуры 2-го ранга, то всегда можно заменить S(*ласкала*)S(*кошку*) на S(*стояла*) (т. е. переходный глагол с дополнением на непереходный глагол). Итак, S-структура 3-го ранга, полученная из данной, имеет вид:

$$S(\text{девочка})S(\text{стояла}).$$

По поводу приведенного определения S-конфигурации надо отметить еще следующее. В нем вовсе не требуется (как это могло показаться на основе анализа примеров), чтобы результирующий S-конфигурации S_{x_1} был единственным.

В языке мы действительно встречаемся со случаем, когда S-конфигурация имеет два результирующих, это все случаи употребления местоимений, например S-конфигурация S(*весьма*)S(*маленькая*) имела кроме S(*маленькая*) еще один результирующий, а именно S(*эта*), где S(*эта*) есть семейство всех местоимений вроде *эта*, *та*.

Еще одно замечание нужно сделать по поводу рангов S-структур. В силу данного нами определения любая S-структура n-го ранга является одновременно S-структурой (n—1)-го ранга.

Введем теперь еще следующее дополнительное ограничение, накладываемое на выделение конфигураций. Дело в том, что может встретиться случай, когда один и тот же элемент S_k образует конфигурацию 1-го ранга с элементом S_i , стоящим вправо от него, и конфигурацию того же ранга с элементом S_j , стоящим влево от него. Для однозначности мы потребуем, чтобы все конфигурации выделялись в строгом порядке, а именно слева направо.

§ 40. Эквивалентность по рангу

Введем теперь новый вид эквивалентности: эквивалентность по рангу. Мы будем говорить, что S_i эквивалентно S_j по рангу l ($S_i \sim_l S_j$), если во всех S-структурах l-го ранга замена S_i на S_j и S_j на S_i переводит отмеченную S-структуру в отмеченную, а неотмеченную в неотмеченную.

Теорема 5.3. Если S_i и S_j — два результирующих конфигурации k -го ранга $\tilde{S}_{(k)}$, то $S_i \tilde{(k)} S_j$.

В силу этой теоремы местоимения эквивалентны по рангу словам, которые они замещают, но не просто эквивалентны. Отношение «быть местоименным элементом к данному семейству» есть отношение S -эквивалентности, лишенное свойства симметрии.

Таким образом в терминах нашей абстрактной системы можно определить местоимения.

§ 41. Норма S -структуры

Введем следующие определения:

Упрощением данной S -структуры называется структура, полученная из данной S -структуры или ее упрощений путем замены конфигураций их результирующими (путем сокращения).

Базисным упрощением называется упрощение, не содержащее никаких конфигураций. Базисным элементом данной S -структуры называется элемент, входящий в ее базисное упрощение.

Количество базисных элементов данной S -структуры называется ее **нормой**. Множественность результирующих в конфигурациях может создать впечатление, что норма — число переменное, зависящее от того, какие результирующие входили в промежуточные упрощения. Опыт показывает, что в реальных языках это не так, и норма данной S -структуры — число постоянное. Доказать или опровергнуть соответствующее утверждение для абстрактных S -структур пока не удалось. Найдено лишь следующее условие однозначности нормы:

Теорема 5.4. Для того чтобы S -структура имела однозначно определенную норму, достаточно, чтобы в процессе сокращения ранг сокращенных конфигураций не превышал ранга вновь возникающих конфигураций.

Мы будем называть базисное упрощение некоторой S -структуры **простым**, если оно не содержит внутри себя базисного упрощения какой-либо другой S -структуры.

Теорема 5.5. Если в языке ранг отмеченных базисных упрощений ограничен, все базисные упрощения отмеченных S -структур простые и какая-либо отмеченная S -структура имеет норму 1, то и все отмеченные S -структуры данного языка имеют норму 1.

Поскольку первые два условия всегда выполняются в реальных языках, то в сущности теорема говорит о том, что если мы применим нашу модель для анализа некоторого множества фраз, включающего односоставные предложения, например, русского языка, то мы получим противоречий нашей

интуиции результат, что все фразы в данном языке односоставны.

Этот результат можно рассматривать как признак несовершенства нашей модели. Такое заключение в известной степени справедливо. В § 44 и далее мы рассмотрим модель, более совершенную с этой точки зрения.

С другой стороны, однако, односоставные предложения типа *Мороз* или *Морозит* являются исключениями и поэтому вполне допустимо построение моделей, в которых предполагается, что все односоставные фразы не отмечены.

В § 44 мы увидим, что можно предложить определение предложения, близкое к интуитивному и в то же время не противоречащее изложенным здесь результатам.

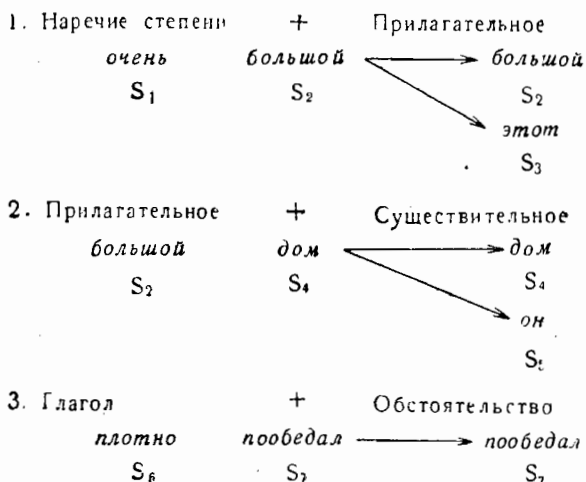
§ 42. Регулярные и нерегулярные конфигурации

Мы будем называть конфигурацию *регулярной*, если:

1) один из результирующих данной конфигурации совпадает с самым правым элементом конфигурации и ни один из результирующих не совпадает с каким-нибудь другим элементом конфигурации; или

2) один из результирующих данной конфигурации совпадает с самым левым ее элементом и ни один из результирующих не совпадает с каким-нибудь другим элементом конфигурации.

Пример. Регулярными конфигурациями в русском языке будут, например:



Ясно, что регулярные конфигурации представляют лингвистической точки зрения наибольший интерес. Поэтому

важно иметь возможность рассматривать некоторые нерегулярные конфигурации как части регулярных.

Введем следующее вспомогательное понятие. Мы будем говорить, что конфигурация является *предбазисной*, если она нерегулярна и ее результирующий есть элемент базисного упрощения. В приложении показан способ сокращения любой S-структуры так, что выделяются лишь регулярные конфигурации и предбазисные конфигурации (теорема 5.6). В дальнейшем изложении предполагается, что сокращение производится именно этим способом.

§ 43. Составляющие и отношения между ними. *Предложение*

Мы будем называть *ядром* тот элемент регулярной конфигурации, который совпадает с результирующим, и *атрибутом* — ту часть регулярной конфигурации, которая остается после вычленения ядра. Если атрибут расположен вправо от ядра, мы будем называть его *правым атрибутом*; аналогично определяется *левый атрибут*.

Назовем *ядро собственным ядром* данного атрибута, если оно является единственным ядром, с которым может встречаться данный атрибут.

Наши «ядро» и «атрибут» вполне соответствуют «ядру» и «спутнику» (satellite) в терминологии Пайка и Питмена⁶. Термин «атрибут» имеет то преимущество, что апеллирует к интуиции лингвиста, напоминая об отношениях в группе существительного. Недостаток его в том, что он, во-первых, имеет более широкое значение, поскольку применяется и к группе глагола; например, по нашей терминологии атрибутом является S(*хорошо*) в конфигурации S(*хорошо*)S(*поет*). С другой стороны, в языках с изафетным построением определение не является атрибутом в нашем смысле, так как конфигурация определяющее имя + определяемое не является регулярной.

Мы не настаиваем на нашей терминологии; нужно лишь подчеркнуть, что самое понятие атрибута (или спутника) является чрезвычайно плодотворным.

В частности, это понятие позволяет решить чисто формально без обращения к смыслу такую проблему, как разграничение «полнозначных» и «служебных» слов (или морфем). *Служебным словом* можно назвать такое слово, которое никогда не имеет атрибутов (спутников). *Полнозначным словом* можно назвать всякое слово, которое не является служебным. Служебные слова можно в свою очередь разбить на

⁶ См., например: P i t t m a n. Nuclear structures in linguistics. «Language», 24, 1949, стр. 275.

факкультативные, а именно те, которые могут быть атрибутами (например, *бы* в русском языке) и обязательными, которые не могут быть атрибутами.

Примечание. Когда мы говорим, что наши термины «ядро» и «атрибут» соответствуют «ядру» и «спутнику» у Пайка и Питмена, то следует иметь в виду, что это соответствие моделирующих терминов терминам конкретного лингвистического описания. Питмен в цитируемой работе приводит десять принципов, по которым конструкция делится на ядро и спутник. Их можно разбить на следующие группы:

1) Логические принципы. Сюда относится, во-первых, избранный нами принцип одинаковости распределения как всей конструкции, так и ее части и, во-вторых, так называемый принцип независимости: А есть ядро в конструкции АВ, если А встречается без В, а В без А не встречается. Этот второй принцип отражен в определении собственного ядра.

2) Статистические принципы. Сюда относится, во-первых, принцип размера класса. А именно, если в какой-нибудь конструкции АВ один из элементов А входит в класс, насчитывающий большее количество членов, чем классы, в которые входят остальные элементы, то этот элемент А считается ядром. Например, в конструкции *буду читать* количество элементов, взаимозаменяемых с *буду* (сюда входят, например, *стану, начну, продолжю*), значительно меньше, чем количество элементов, взаимозаменяемых с *читать*. Поэтому *читать* считается ядром, а *буду* — спутником. Это, так сказать, статистика, основанная на парадигматических отношениях. Другой статистический принцип Питмена основан на синтагматике. Если А из конструкции АВ встречается еще в конструкциях вида АД или ДА, где Д — различные группы слов, а В встречается в конструкциях вида ВС или СВ, причем первых значительно больше, чем вторых, то А считается ядром, а В спутником. Оба эти принципа не учитывались в нашем определении ядра и атрибута. Если на основе нашей логической модели будет построена статистическая, то эти принципы легко могут быть формализованы.

3) Просодические принципы. Сюда относится учет ударения, интонации и т. п.

4) Семантические принципы. Основаны на учете значения. Ясно, что принципы двух последних групп не могут быть отражены в нашей модели, а отсюда следует, что не все, что Питмен делит на ядро и спутник, может быть разделено на ядро и атрибут в терминах нашей модели.

В дескриптивной лингвистике мы встречаемся с чрезвычайно важным понятием *составляющего*. Постараемся определить теперь соответствующее понятие в терминах нашей системы. Мы будем называть составляющими данной S-структуры:

а) любой элемент исходной S-структуры, входящий в базисное упрощение; такой составляющий называется *собственным базисным составляющим*;

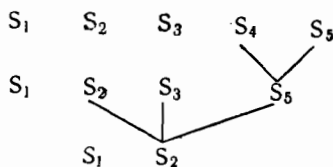
б) любую предбазисную конфигурацию исходной S-структуры; такой составляющий называется *несобственным базисным составляющим*;

в) любой атрибут, элементы которого входили в исходную S-структуру; такой составляющий называется атрибутивным.

При нашем способе сокращения можно сделать так, чтобы любой элемент исходной S-структуры входил в один и только один составляющий.

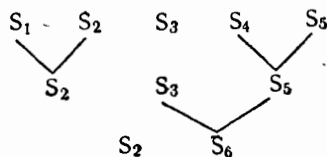
Примеры:

1) Возьмем фразу *Он поднялся на высокий холм* и соответствующую ей S-структуру $S_1S_2S_3S_4S_5$. Сокращение ее можно представить следующей схемой:



Таким образом, мы имеем следующие составляющие: два собственно базисных составляющих — $R_1 = S_1$ (*он*), $R_2 = S_2$ (*поднялся*) и два атрибутивных составляющих — $R_3 = S_4$ (*высокий*), $R_4 = S_3S_5$ (*на холм*).

2) Возьмем фразу *Советские шахматисты одержали крупную победу* и соответствующую ей S-структуру $S_1S_2S_3S_4S_5$. Сокращение ее можно представить следующей схемой:



Здесь S_6 — семейство, объединяющее только непереходные глаголы типа *уехали*. (Мы исходим из того, что фраза типа «шахматисты одержали» не является отмеченной, в то время как фраза «шахматисты уехали» отмечена.)

В этой структуре можно выделить следующие составляющие: $R_1 = S_1$ (*советские*), $R_2 = S_2$ (*шахматисты*), $R_3 = S_4$ (*крупную*), $R_4 = S_3S_5$ (*одержали победу*). S_3S_5 — пример предбазисного составляющего. S_6 не есть составляющий, так как не входил в исходную структуру.

Введем понятие непосредственного подчинения между составляющими данной структуры.

Если некоторый составляющий А является атрибутом к ядру С, а С входит в качестве составной части в В, или совпадает с В, то мы будем говорить, что А непосредственно под-

ч и н е н $B (B \Rightarrow A)$. Если два составляющих A и B являются собственными или несобственными базисными составляющими, то мы будем говорить, что A непосредственно подчинен B , а B непосредственно подчинено $A (A \Leftarrow B)$. В этом случае мы будем говорить, что A и B составляют п р е д и к а т и в н у ю п а р у.

П р и м е р. В первом примере имеем $S_1 \Leftarrow S_2, S_2 \Rightarrow S_3S_5, S_3S_5 \Rightarrow S_4$. Во втором примере $S_2 \Leftarrow S_3S_5, S_3S_5 \Rightarrow S_4, S_2 \Rightarrow S_1$.

Мы будем говорить, что составляющий A подчинен составляющему B , если имеется цепочка составляющих R_1, R_2, \dots, R_n , такая, что $R_1 = A, R_n = B$ и $R_{i+1} \Rightarrow R_i (1 \leq i < n)$.

Введем теперь следующее определение:

S -структура называется п р е д л о ж е н и е м, если для любого составляющего R_i имеется один и только один составляющий R_j , такой, что $R_j \Rightarrow R_i (i \neq j)$.

Почти очевидны следующие утверждения.

Теорема 5.7. Для любых двух составляющих R_1 и R_2 данного предложения имеет место одно из двух: 1) один из них непосредственно подчиняет другой; 2) имеется составляющий R_3 , который подчиняет R_2 и одновременно подчиняет R_1 .

Теорема 5.8. В предложении имеется одна и только одна предикативная пара.

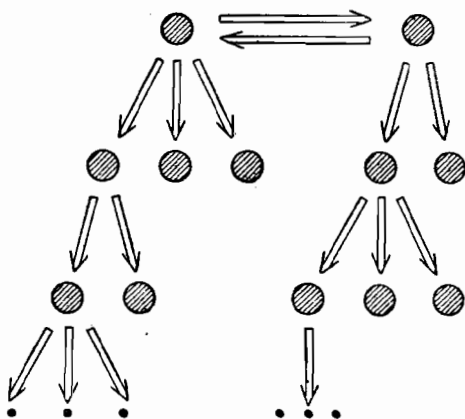
С л е д с т в и е. Предложение имеет норму 2.

Интересно, что верно и обратное утверждение.

Теорема 5.9. Если отмеченная S -структура имеет норму 2, то она является предложением.

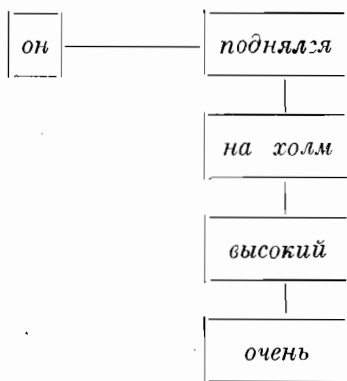
Таким образом, имеется другое, более удобное определение предложения:

Предложение есть S -структура нормы 2 (интересно сопоставить это определение с теоремой 5.5).



В этом определении понятие подчинения совершенно не используется. Однако это понятие имеет определенный смысл, а именно: оно позволяет формальным путем строить схемы подчинения, почти аналогичные тем, которые разбираются в школьной грамматике. Мы знаем теперь, что общая схема предложения имеет вид, указанный на схеме стр. 131.

Легко проверить, что в эту общую схему почти всегда умещаются школьные схемы; например, для фразы *Он поднялся на очень высокий холм*:



Для фразы *Советские шахматисты одержали крупную победу* единственное отличие нашей схемы от школьной состоит в подходе к группе предиката:

Наша схема

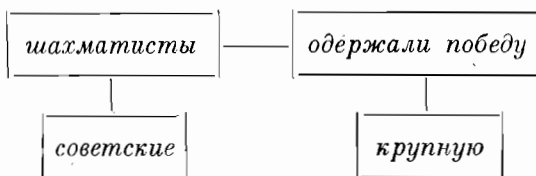
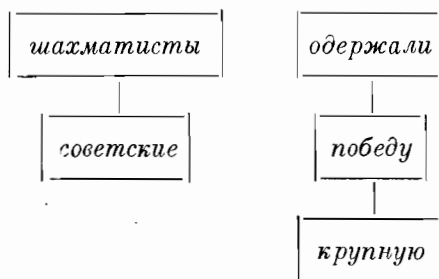


Схема школьной грамматики



Имеется, однако, существенное различие между нашими схемами и схемами школьной грамматики с точки зрения того, между какими элементами устанавливаются отношения. Мы устанавливаем отношения между составляющими, а традиционная грамматика устанавливает отношения между членами предложения, т. е. тем, что в наших терминах можно определить как некоторые классы составляющих.

Но мы можем подобные классы определить формально в терминах нашей модели, воспользовавшись аналогией с понятием структурного объединения фонем, упомянутым в § 15.

Назовем *р е л я т е м о й* класс составляющих, каждый из которых хотя бы в одной S-структуре подчинен данному фиксированному составляющему, не подчиняя его, и заменим на схеме отдельные составляющие релятемами. Тогда мы получим некоторое приближение к понятию члена предложения.

Удобство понятия релятемы можно продемонстрировать на следующем примере. Для S-структур предложений «Красивая дочь вошла» и «Дочь рыбака вошла» нужны две разные схемы, если элементами считать составляющие. В то же время *рыбак* и *красивая* относятся к одной релятеме. Поэтому этим двум схемам составляющих соответствует общая схема на уровне релятем.

Вообще говоря, релятемы суть пересекающиеся классы составляющих. Однако если все ядра в регулярных конфигурациях собственные и имеется только две релятемы базисных составляющих, то разбиение на релятемы есть разбиение на непесекающиеся классы.

В заключение остановимся на одном формальном способе определения предложения, выдвинутом недавно Бар-Хиллелом⁷. Суть его состоит в следующем. Предложение следует определить рекурсивно, а именно указать некоторый список исходных фраз, которые считаются предложениями, а затем указывается ряд правил вида: Если Z есть предложение, то некоторая фраза Z', полученная применением такого-то преобразования из нее, также есть предложение.

В сущности к этой форме легко можно привести наше определение, а именно задать некоторое множество структур, состоящих из двух собственных базисных составляющих, а затем сказать, что предложением является любая отмеченная S-структура, которую можно путем сокращения привести к данному виду.

Если, однако, в исходной список включить и структуры нормы 1, то возникнут трудности, охарактеризованные в теореме 5.5.

⁷ J. Bar-Hillel. On recursive definitions in empirical sciences, «Proceedings of the XI-th International Congress of Philosophy», vol. 5, 1953, стр. 163.

§ 44. Конструкции

В предыдущих параграфах наша модель основывалась на сокращении конфигураций и получении на этом основании некоторых базисных структур (назовем ее первой моделью). Теперь рассмотрим новую (вторую) модель, отличающуюся от предыдущей тем, что в ней сокращению подлежат только регулярные конфигурации, т. е. сокращение есть вычеркивание некоторых атрибутов. Базисное упрощение, полученное при помощи такого сокращения, мы будем называть конструкцией (конструкция совпадает с базисом в том случае, если ее элементы не входят в некоторые нерегулярные конфигурации). S-структура называется подконструкцией данной конструкции, если она приводится к данной конструкции путем вычеркивания некоторых атрибутов.

Важность этой, второй модели синтаксического описания в том, что реальная грамматика лучше всего моделируется при помощи модели, представляющей комбинацию первой и второй моделей.

В частности, конструкции, выделенные во второй модели, соответствуют так называемым «логико-грамматическим типам предложения»⁸.

Так, для русского языка можно выделить, например, такие конструкции (для простоты эквивалентные по рангу семейства объединяются, а вместо семейств выписываются отдельные представители):

- I. *Мальчик идет*
Мальчик будет ходить
- II. *Мальчик берет книгу*
Мальчик будет брать книгу
- III. *Мальчик красивый*
Мальчик кажется красивым
- IV. *Мальчик — поэт*
Мальчик будет поэтом
Мальчик хочет быть поэтом
- V. *Мальчик рожден богиней*
Мальчик рождается богиней
- VI. *Книга принадлежит мальчику*
Книга будет принадлежать мальчику
- VII. *Мальчик называет книгу «никой»*
Мальчик будет называть книгу «никой»

⁸ См.: В. Г. Адмони. Введение в синтаксис современного немецкого языка. М., 1955, стр. 102 и сл.

Ясно, что в этой модели теорема 5.5 не имеет места, и поэтому при определении предложения необязательно требовать, чтобы базисная структура имела норму 2. Можно дать такое определение:

Предложением называется отмеченная S -структура, в которой для любых составляющих R_1 и R_2 имеет место одно из двух:

- 1) один из них подчиняет другой;
- 2) имеется составляющий R_3 , который подчиняет R_2 и одновременно подчиняет R_1 .

Сравнение этого определения с теоремой 5.7 показывает, что всякая S -структура, являющаяся предложением по определению, данному в § 43, есть предложение и в смысле определения этого параграфа. Обратное, разумеется, не всегда имеет место.

§ 45. Должно ли множество отмеченных фраз быть конечным или бесконечным?

В § 2 мы оставили открытым этот вопрос. Теперь мы должны вернуться к нему. Весьма целесообразной как для теоретических исследований, так и для практических использований была бы следующая концепция.

Множество отмеченных фраз конечно. Это, например, совокупность фраз данной книги, или всех книг данного писателя, или, например, всех книг польских писателей XVI в. (как показывает работа польских ученых над частотным словарем польского языка XVI в., это множество вполне обозримо). При этом, конечно, придется считать все не встретившиеся фразы неотмеченными, так что неотмеченных фраз может получиться уже бесконечное множество, однако к неотмеченным фразам можно вообще не обращаться.

Оказывается, что подход к множеству отмеченных фраз как конечному заставляет отказаться от такого важного понятия как «регулярная конфигурация». Имеет место следующее утверждение.

Теорема 5.10. Если язык задан как конечное множество отмеченных фраз, то в нем не может быть выделено ни одной регулярной конфигурации.

Таким образом, множество отмеченных фраз в теоретико-множественной концепции О. С. Кулагиной должно мыслиться как бесконечное. Итак, мы пользуемся абстракцией актуальной бесконечности, т. е. оперируем сразу всем бесконечным множеством фраз. Как справедливо заметил А. А. Марков в своем выступлении на конференции по машинному переводу, такой подход существенно осложняет все наши операции. В самом деле, как убедиться в том, что какие-то элементы

E_i и E_j действительно эквивалентны. Очевидно, мы должны перебрать все отмеченные фразы одну за другой (а именно так и должна была бы поступать машина), где встречаются эти элементы, но мы не можем быть уверены, что этот процесс перебора фраз когда-либо остановится.

Поэтому целесообразно выяснить, каким образом следовало бы изменить изложенную в главе III систему исходных понятий, с тем чтобы она была применима к конечному множеству отмеченных фраз, и выяснить следствия для ее лингвистической интерпретации, вытекающие из такого видоизменения.

1. Первый шаг состоит в следующем. В определении В-эквивалентности требовалось, чтобы два элемента V_i и V_j были взаимозаменяемыми в любой произвольной структуре. Мы можем теперь ослабить это требование, проверяя вхождения V_i и V_j не во все произвольные структуры, а во все структуры длины, не превосходящей некоторое число⁹ k . Определение В-эквивалентности примет тогда следующий вид:

V_i ограничено эквивалентно V_j , если из отмеченности любой В-структуры длины, не превосходящей k и вида $V(A_1)V_jV(A_2)$ следует отмеченность В-структуры $V(A_1)V_iV(A_2)$ и из отмеченности любой В-структуры длины, не превосходящей k и вида $V(D_1)V_jV(D_2)$ следует отмеченность В-структуры $V(D_1)V_iV(D_2)$.

Аналогично определяется конфигурация и эквивалентность по рангу.

Идея ограничения длины окружений, в которых проверяется взаимозаменяемость элементов, принадлежит Харрису, хотя она и связана у него с другими построениями.

Заметим теперь, что наше видоизменение основных определений не затрагивает тех моделей, которые с их помощью построены, ибо легко проверить, что ограниченная эквивалентность обладает свойствами симметрии, рефлексивности и транзитивности. Между тем смутившая нас теорема 5.10 в этих условиях уже не имеет места.

Если теперь количество слов конечно, то и количество фраз заданной длины конечно, и анализ текста с помощью моделей этой главы может мыслиться как практически осуществимый.

Это, однако, не означает, что практически может быть осуществлен алгоритм машинного анализа текста, который работал бы на основе даже видоизмененного определения. Дело в том, что если иметь в виду не строгое понятие эквивалентности, необходимое для моделирования, а его лингвистический аналог

⁹ В терминах математической логики можно сказать, что определение В-эквивалентности предлагается строить с использованием оператора «наибольшее число» (см.: В. А. Успенский. Лекции о вычислимых функциях. М., 1960, стр. 85).

(взаимозаменяемость двух элементов или элемента и группы элементов), то, по-видимому, можно и не проверять все вхождения данного элемента во все фразы (пусть даже заданной длины).

Вообще говоря, на конечном множестве фраз при столь сильных требованиях может не оказаться ни одной пары эквивалентных элементов.

Мы уже указывали (§ 23), что в дескриптивной лингвистике рассматривается случай взаимозаменяемости двух элементов хотя бы для одной фразы, и говорили, что при этом мы не гарантированы от всевозможных случайностей (ср. «он сказал, что болен» и «он сказал, директор болен», где что взаимозаменяемо с директор). Это обстоятельство заставляет дескриптивистов прибегать к целой системе оговорок (например, необходимость учета интонации и т. п.). Однако элемент случайности может быть исключен, если мы потребуем, чтобы для достаточно большого числа фраз два элемента были взаимозаменяемы.

2. Мы будем говорить, что V_1 практически эквивалентно V_2 , если существует N фраз A_1 и A_2 , таких, что из отмеченности $V(A_1)V_1V(A_2)$ следует отмеченность $V(A_1)V_2V(A_2)$, и существует N фраз C_1 и C_2 , таких, что из отмеченности $V(C_1)V_1V(C_2)$ следует отмеченность $V(C_1)V_2V(C_2)$.

Итак, мы вновь сталкиваемся с тем же явлением, что и в § 31: при переходе от модели к ее интерпретации возникает необходимость перейти от логических критериев к статистическим.

Возвращаясь от интерпретации к самой модели, отметим, что рассмотрение множества отмеченных фраз как бесконечно го вовсе не бессмысленно.

При таком подходе мы смогли выяснить все те понятия, которые необходимы для построения синтезирующей модели, обратной по отношению к конфигурационному анализу. Это важно и по следующим соображениям. В § 37—38 мы говорили о неадекватности построенной там модели фактам конкретных языков. Однако сравнить соответствующую модель с конфигурационной удобнее всего, если их рассматривать с точки зрения синтеза (ср. § 3).

Итак, рассмотрим синтезирующую модель, соответствующую конфигурационной.

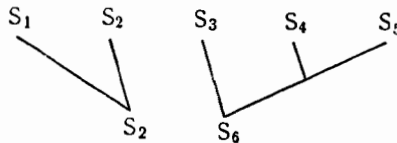
§ 46. Порождение методом развертывания

Используя метод анализа по непосредственно составляющим, Хомский предложил построить более сильный порождающий процесс, чем язык с конечным числом состояний. Как мы видели, конфигурационный анализ является дальнейшей формализацией метода непосредственно составляющих, поэтому

мы и будем вести, в отличие от Хомского, все изложение в терминах конфигурационного анализа.

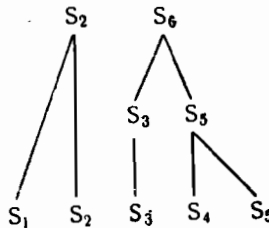
Вспомним второй пример из § 43. Там мы исходили из всего множества фраз, и путем анализа получили следующую схему сокращения:

Советские (S_1) *шахматисты* (S_2) *одержали* (S_3) *крупную* (S_4) *победу* (S_5)



Пусть теперь нам, наоборот, известно, что S_2S_6 есть некоторая базисная структура, являющаяся предложением, пусть известны конфигурации S_1S_2 , S_4S_5 , S_3S_5 , причем известно, что первая разворачивается из S_2 , вторая — из S_5 , а третья — из S_6 .

Тогда весь процесс получения данной фразы представится схемой или, как иногда говорят, деревом:



Пусть каждая базисная структура обозначается символом Z . Тогда процесс получения нашей фразы можно представить как последовательность операций:

- 1) $Z \rightarrow S_2S_6$
- 2) $S_2 \rightarrow S_1S_2$
- 3) $S_6 \rightarrow S_3S_5$
- 4) $S_5 \rightarrow S_4S_5$
- 5) $S_1 \rightarrow \text{советские}$
- 6) $S_2 \rightarrow \text{шахматисты}$
- 7) $S_3 \rightarrow \text{одержали}$
- 8) $S_4 \rightarrow \text{крупную}$
- 9) $S_5 \rightarrow \text{победу}$

Все эти замены суть операции типа $x \rightarrow y$, причем в левой части обязательно стоит один символ, а в правой части может быть несколько символов. Каждый ряд операций типа $x \rightarrow y$ приводит к преобразованию исходного символа, которое можно записать следующим образом:

Z	
S_2S_6	(по 1)
$S_1S_2S_6$	(по 2)
$S_1S_2S_3S_5$	(по 3)
$S_1S_2S_3S_4S_5$	(по 4)
<i>советские</i> $S_2S_3S_4S_5$	(по 5)
.	

Советские шахматисты одержали крупную победу (по 9).

Такой ряд мы будем называть в ы в о д о м (derivation). Символы Z и S мы будем называть мета-обозначениями. Строку вывода, не содержащую никаких мета-обозначений, назовем з а к л ю ч и т е л ь н о й строкой. Совокупность заключительных строк, задаваемых правилами данного типа, будем считать множеством отмеченных фраз, порождаемым данной грамматикой.

При этом грамматика в общем виде задается некоторой совокупностью базисных структур: Z_1, Z_2, \dots, Z_n и некоторым набором правил вывода F , которые имеют вид:

$$\begin{aligned}
 x_1 &\rightarrow y_1 \\
 x_2 &\rightarrow y_2 \\
 &\dots \\
 x_n &\rightarrow y_n,
 \end{aligned}$$

где каждая левая часть содержит ровно один элемент. Набор отмеченных фраз, т. е. заключительных строк вывода, Хомский назвал т е р м и н а л ь н ы м я з ы к о м (terminal language).

Теорема 5.11 (Хомский). Всякий язык с конечным числом состояний есть терминальный язык, но не всякий терминальный язык есть язык с конечным числом состояний.

Таким образом, модель терминального языка обладает большей общностью и эффективностью, чем модель языка с конечным числом состояний.

Эта модель интересна и с кибернетической точки зрения. Именно на ее основе была высказана очень интересная гипотеза, которую мы рассмотрим в следующем параграфе.

§ 47. Гипотеза Ингве

Недавно крупный американский специалист в области машинного перевода Ингве высказал следующую интересную гипотезу¹⁰.

Предположим, что человек действительно синтезирует фразу рассмотренным выше способом. Ясно, что он не выдает сразу готовую фразу, а выстраивает слова цепочкой одно за другим, при этом он не всегда знает, как будет закончена фраза, начатая определенными словами. Фраза может быть сколь угодно длинной (примером может служить шутливое стихотворение «Дом, который построил Джек»). С другой стороны, объем быстродействующей памяти человека, как показали опыты Миллера¹¹, весьма ограничен. Ингве предлагает модель, которая реализует эту ситуацию, а именно машину, состоящую из следующих частей:



При этом в постоянной памяти содержатся правила вида $x_i \rightarrow y_i$, в решающем устройстве производится сама операция замены, а в быстродействующей памяти хранятся те промежуточные сведения, которые необходимо запомнить для выполнения некоторых операций.

Например, для вывода последовательности слов *очень сильная слониха ударила льва* по правилам:

- 1) *слониха* \rightarrow *сильная слониха*
- 2) *сильная* \rightarrow *очень сильная* и т. д.

до того, пока не сработало второе правило, в быстродействующей памяти надо хранить слова *сильная слониха*, которые могут быть выданы (и стерты в быстродействующей памяти) лишь тогда, когда выведено слово *очень*.

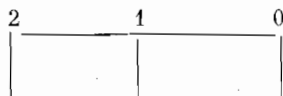
Из этого примера ясно, что быстродействующая память должна вмещать по крайней мере два слова, чтобы произвести нужную нам последовательность слов.

¹⁰ V. H. Y n g v e. A model and a hypothesis for Language Structure, «Proceedings of American philosophical Society», vol. 104, N 5, 1960.

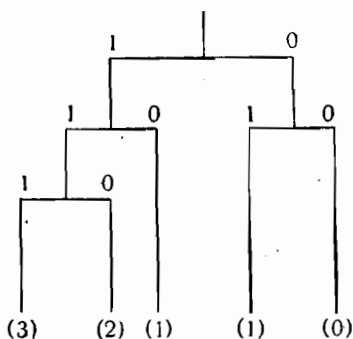
¹¹ G. A. M i l l e r. Human memory and the storage of information. «IRE Transactions on Information Theory», vol. IT-2, 1956, N 3.

Ингве предложил остроумный метод определения максимального объема памяти, необходимого для производства некоторого предложения, по виду «дерева» этого предложения.

Для этого применяется следующий прием. В каждой точке, где происходит ветвление дерева, ветви нумеруются справа налево, причем самая правая ветвь получает индекс 0, следующая — индекс 1 и т. д., например:



Для приведенной фразы мы получим дерево со следующими индексами:



Далее подсчитывается сумма индексов, стоящих на ветви, идущей от некоторого слова до вершины дерева. Соответствующие суммы проставлены снизу.

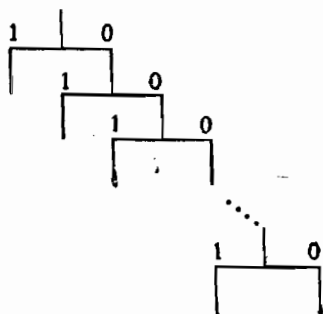
Наибольшая из этих сумм называется глубиной фразы. Максимальная глубина фразы в данном языке называется глубиной языка.

Ингве показывает, что глубина языка как раз характеризует минимальный объем быстродействующей памяти, необходимый для порождения описанным устройством любой фразы данного языка.

Одним из интереснейших следствий этого факта является разница между так называемыми прогрессивными и регрессивными¹² конструкциями в языке.

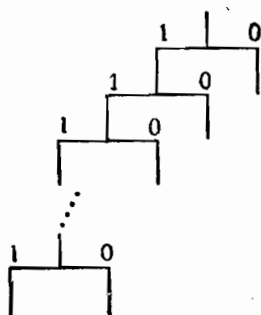
¹² Следует подчеркнуть, что термины «прогрессивная» и «регрессивная» структура употребляются здесь в гораздо более узком смысле, чем, например, прогрессивная и регрессивная последовательность в кн.: Ш. Б а л л и. Общая лингвистика и вопросы французского языка. Перевод с франц. М., 1955, стр. 218—312.

Прогрессивными конструкциями Ингве называет деревья или «поддеревья» (части деревьев), имеющие вид:



т. е. такие, в которых ветвление происходит в точках с нулевым индексом. Ингве показывает, что указанное устройство при ограниченном объеме быстродействующей памяти может производить фразы сколь угодно большой длины, если деревья этих фраз суть прогрессивные конструкции.

Регрессивными конструкциями Ингве называет деревья или поддеревья, имеющие вид:



т. е. такие, где ветвление происходит в точках с ненулевым индексом. Именно за счет таких конструкций и растет глубина фразы. После этих предварительных замечаний перейдем к изложению основной гипотезы Ингве:

1) Фразы, действительно употребляемые в разговорном языке, имеют глубину, не превосходящую определенного небольшого числа (например, 7 ± 2 ; таков, по Миллеру, объем быстродействующей человеческой памяти).

2) Во всех языках существуют методы ограничения регрессивных конструкций, а при наличии фраз с большой глубиной используются синонимичные фразы меньшей глубины.

3) При изменении языка фактор глубины должен играть существенную роль.

Первое утверждение проверить достаточно просто. Что касается второго и третьего, то здесь необходимо прежде всего выяснить, какими способами можно ограничить регрессивные конструкции. Из всех методов, исследуемых в этой связи Ингве, мы остановимся на следующих: ограничение ранга, структурная перестройка, предпочтение бинарным конструкциям, особенно для базиса и в точках ветвления.

Ограничение ранга конфигураций характерно для регрессивных конструкций, а именно для определительных групп. Ингве справедливо указывает, что нет особых семантических оснований для разграничения существительных и прилагательных (в ряде языков, например в болгарском или в финно-угорских, некоторые существительные имеют даже степени сравнения), прилагательных и наречий. Если все же эти категории, как правило, выделяются, то не следует ли — спрашивает Ингве — искать причину в том, что появление наречия, например *очень* в «*очень* красивая девушка пела», служит сигналом того, что глубину фразы, достигшую величины 3, нельзя больше увеличивать?

Только широкие типологические исследования могут помочь опровержению или подтверждению этой части гипотезы.

Отметим лишь, что в языках, допускающих препозицию несогласованного определения или даже требующих ее, по-видимому, нет никаких структурных сигналов о достижении определенной глубины, и глубина контролируется лишь стилистическими соображениями. Сам Ингве приводит пример:

His mothers brothers sons daughters hat
'Его' 'матери' 'брата' 'сына' 'дочери' 'шляпа',

где глубина достигает 5.

Если, однако, в английском языке препозитивная конструкция с *'s* употребляется гораздо реже, чем постпозитивное определение с *of*, то в ряде других языков, например эстонском, литовском, армянском и др., препозиция несогласованного определения является правилом. Поэтому интересно было бы исследовать вопрос о том, является ли глубина в таких группах соответствующего языка структурно ограниченной или нет.

Структурная перестройка, при которой уменьшается глубина за счет замены регрессивного построения прогрессивным — факт, подтверждаемый материалом ряда языков. Например, в русском языке когда-то были возможны группы с препозицией склоняемой сравнительной степени, например:

огромнейший первого камень схватил («Одиссея» в переводе Жуковского);

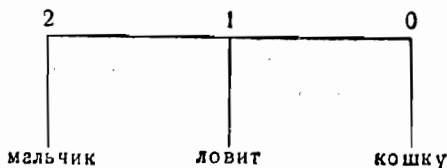
вскричала она в *еще сильнейшем* испуге (Достоевский. Униженные и оскорбленные).

Сейчас в таких группах возможна только постпозиция, что уменьшает глубину фразы.

В общем случае, однако, трудно утверждать, что в истории языка всякое изменение шло в сторону сокращения глубины.

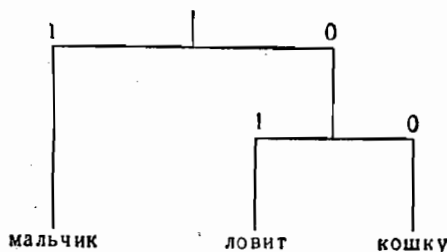
Предпочтение бинарным конструкциям, пожалуй, самая характерная черта языка.

Вообще говоря, неясно, почему предложение *Мальчик ловит кошку* не описывается обычно в грамматике деревом:



Это было бы вполне адекватным выражением двухместного отношения R (*мальчик, кошка*).

Предпочтение дерева:



объяснимо только с точки зрения гипотезы Ингве.

Между прочим, гипотеза Ингве вполне согласуется с тем предположением, что базисная структура не должна содержать больше двух элементов (ср. § 44).

Общий лингвистический интерес гипотезы Ингве в том, что здесь впервые модель строится с учетом того, что язык создается не произвольным устройством, а именно человеком, и поэтому она должна отражать специфические особенности именно данного кибернетического устройства.

Мы видим, таким образом, что модель терминального языка имеет глубокий лингвистический смысл и позволяет ставить ряд новых вопросов.

Однако оказалось — и доказательство этого факта является основной заслугой Хомского, — что эта модель недостаточна для описания важнейших сторон языка.

Есть целый ряд явлений в языке, которые плохо описываются данной моделью. Прежде всего требование замены всего одного символа не дает возможности породить очень простые фразы, например, *Вчера я был болен* нельзя получить этим способом из *я болен*.

Во-вторых, порождение фраз типа *Дом строится рабочими* из фраз типа *Дом стоит* противоречит нашему интуитивному восприятию. Мы бы хотели иметь такую модель, в которой фраза *Дом строится рабочими* порождалась из фразы *Рабочие строят дом*.

Ясно, однако, что для создания модели порождения, отражающей все эти явления, должна существовать соответствующая аналитическая модель. Между тем, в предыдущих аналитических моделях не могли быть отражены все эти явления.

§ 48. Трансформационный анализ

В аналитической модели конфигурационного анализа мы рассматривали лишь преобразования, состоящие в замене конфигураций (произвольных — в первой модели и только регулярных — во второй модели) их результирующими. Введем теперь новый вид преобразований, которые мы назовем *т р а н с ф о р м а ц и я м и*. При этом трансформацией мы будем называть преобразования, совершаемые не над S-структурами, но над конкретными фразами; однако при этом будет учитываться характер соответствующей S-структуры. После этих предварительных замечаний мы перейдем к основным определениям. Мы будем говорить, что две отмеченные фразы *A* и *B* находятся друг к другу в *о т н о ш е н и и т р а н с ф о р м а ц и и* или *т р а н с ф о р м и р у ю т с я* друг в друга, если выполняются следующие два условия:

1. а) Хотя бы одно слово в *A* входит в неединичную окрестность (т. е. окрестность, состоящую более чем из одного слова). б) Для любого слова *x* в *A*, входящего в неединичную окрестность, найдется в *B* слово *y*, входящее в $\Gamma(x)$ и обратно для любого слова *y* в *B*, входящего в неединичную окрестность, найдется слово *x*, входящее в $\Gamma(y)$.

2. Для любой фразы A_1 , имеющей ту же S-структуру, что и *A*, найдется фраза B_1 , имеющая ту же S-структуру, что и *B* и связанная с A_1 с первым условием. Аналогичное требование предъявляется к любой фразе, имеющей ту же S-структуру, что и *B*.

Легко проверить, что отношение трансформации разбивает все множество отмеченных фраз на непересекающиеся классы, поскольку оно обладает свойствами рефлексивности (каждая фраза находится в отношении трансформации к самой себе), симметрии (если *A* трансформируется в *B*, то *B* трансформируется

в А) и транзитивности (если А трансформируется в В, а В трансформируется в С, то А трансформируется в С).

В каждом из получившихся классов фраз мы выделим фразу, относящуюся к кратчайшей конструкции (при интерпретации это будет конструкция, анализ которой средствами первой модели наименее противоречит нашему интуитивному восприятию) и назовем ее фразой ядра.

Набор фраз ядра составляет синтаксическое ядро языка.

Первое требование мы назовем требованием семантической связи, поскольку, как мы знаем, слова, входящие в одну окрестность, — это, при интерпретации, разные формы одного слова¹³. Слова, входящие в единичную окрестность, — это, как правило, служебные слова (во всяком случае верно, что каждое служебное слово имеет единичную окрестность), например предлоги. Вполне разумно разрешить, чтобы при трансформации предлоги исчезали, появлялись или один заменялся другим. Наша модель должна быть, по-видимому, пересмотрена в применении к «аморфным языкам», но для наших целей это несущественно.

Второе требование назовем требованием совершенства трансформации. Надо сказать, что это требование отражает важную черту преобразований, рассматриваемых в лингвистике и получивших в последнее время название «трансформаций», а именно регулярность некоторого преобразования, его применимость к целому множеству фраз. В модели, разумеется, подобное требование должно быть очень строгим. Ясно, что при интерпретации предложенной модели мы столкнемся в каждом реальном языке с рядом исключений. Поэтому введем следующие определения. Пусть на некотором подмножестве множества отмеченных фраз определена трансформация, т. е. преобразование, отвечающее выдвинутому выше условию. Если это преобразование остается трансформацией на всем множестве отмеченных фраз, представляющем некоторый конкретный язык, то мы будем называть этот язык языком с полной трансформацией (сокращенно — ЯПТ). Если же трансформация может быть определена лишь на некотором подмножестве отмеченных фраз, т. е. лишь на части фраз некоторого конкретного языка, то мы будем называть этот язык языком с неполной трансформацией.

Возьмем, например, эстонский язык, для которого характерно следующее:

¹³ Пункт «б» первого условия соответствует условию трансформации, выдвинутому в работе: В. В. Бородин. К модели описания языка. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста», вып. 6. М., 1961, стр. 5.

1) от любого глагола можно образовать отглагольное существительное (nomen actionis) с суффиксом *-mine*, например:

kirjutama 'писать' *kirjutamine* 'писание'
lugema 'читать' *lugemine* 'чтение'
töötama 'работать' *töötamine* 'работа' и т. д.

2) от любого глагола можно образовать nomen agentis с суффиксом *-ja*, например:

kirjutaja 'писатель', 'пишущий'
lugeja 'читатель'
töötaja 'трудящийся'

3) от любого глагола (как переходного, так и непереходного) можно образовать причастие не только действительного залога на *-nud*, например:

kirjutanud 'пишущий'
lugenud 'читающий'
töötanud 'работающий',

но и причастие страдательного залога на *-tud*:

kirjutatud
lugetud
töötatud

4) от любого глагола образуются две формы инфинитива и в том числе инессивная форма второго инфинитива, соответствующая русскому деепричастию, например:

kirjutades русск. *пиша* неупотребительно
lugedes 'читая' и т. п.

Поэтому, если имеется фраза:

Õpilane loeb raamatut 'Ученик читает книгу',

то возможны все трансформации:

raamatu lugemine 'чтение книги'
raamatu lugeja 'читатель книги'
raamatut lugenud õpilane 'читающий книгу ученик' и т. п.

Можно показать, что в каждом языке можно выделить часть фраз, такую, что эта часть является ЯПТ. Важным типологическим критерием является разный объем этой части в различных языках. Например, сравнивая с этой точки зрения различные славянские языки, мы можем установить, что польский язык, по-видимому, ближе к идеальному ЯПТ, чем чешский, а чешский ближе, чем русский.

В польском языке от каждого глагола можно образовать имя действия с суффиксом *-anie/-enie*. Если при этом глагол имел при себе показатель возвратности *się*, то соответствующий показатель остается и при отглагольном имени, например: *Co się dzieje podczas uczenia się drugiego języka i posługowania się nim?* Существительному свойственны формы вида, ср. *uczenie — nauczanie*. Все это расширяет возможности трансформации.

В чешском языке употребление возвратной морфемы при трансформации иногда возможно (главным образом для ясности), например: *učení se cizímu jazyku* 'обучение самого себя иностранному языку', но *učení cizímu jazyku* 'обучение другого'. Однако такие случаи все же исключения. Зато, как известно, видовые различия последовательно сохраняются, ср. *dělení — rozdělení* и т. п. Эти примеры показывают, что и понятие ЯПТ, вернее близости к идеальному ЯПТ, достаточно интересно с точки зрения типологии славянских языков.

По-видимому, это понятие может оказаться важным и для диахронического изучения языка. Можно высказать гипотезу, что развитие каждого языка идет в направлении к ЯПТ. Не обладая соответствующими данными из области славянских языков, мы приведем аргумент в пользу этой гипотезы, взятый из немецкого языка. В немецком языке почти от каждого переходного глагола можно образовать имя действия с суффиксом *-ung* и тем самым получить трансформированную конструкцию для любой фразы ядра, содержащей переходный глагол. Исключение составляет группа наиболее древних и наиболее простых переходных глаголов без префиксов, например *machen* 'делать' (нет *Machung*, но есть *Abmachung, Vermachung, tragen* 'нести' (нет *Tragung*, но есть *Abtragung, Austragung*) и т. п.

Связано это с тем, что имена на *-ung* лишь постепенно были втянуты в сферу трансформированных конструкций. Число *nomina actionis* с суффиксом *-ung* постепенно росло¹⁴, но этот процесс не закончился еще до сих пор.

Если гипотеза о диахронической тенденции в направлении к ЯПТ верна, то представляется в какой-то мере неслучайным, что именно финно-угорские языки, в частности эстонский, являются ЯПТ. Понятие ЯПТ, может быть, окажется связанным (конечно, косвенно и многократно опосредствовано) с типологическими признаками, характеризующими, по Трубецкому¹⁵, переход от индоевропейского состояния к алтайскому, а именно с принципом агглютинации, дающим возможность широкого, а главное симметричного образования производных слов.

¹⁴ См.: Н. P a u l. Deutsche Grammatik, Bd V. Halle a. d. S., 1957, 3 Aufl., стр. 74.

¹⁵ См.: Н. С. Т р у б е ц к о й. Мысли об индоевропейской проблеме. — ВЯ, 1958, № 1, стр. 77.

Вернемся к обсуждению модели. Определив трансформацию через термины «S-структура» и «окрестность», мы можем выработать некоторые признаки трансформируемости. Пусть нам известно, что для некоторого преобразования выполняется требование семантической связи, но мы не знаем, выполняется ли требование совершенности. Ответ на данный вопрос дает следующее утверждение.

Теорема 5.12. Если в простом и однородном языке некоторое преобразование удовлетворяет требованию семантической связи, то оно есть трансформация.

Эта теорема показывает, что всякое разумное преобразование, определенное на фразах однородного простого языка, есть трансформация. Иначе говоря, понятия ЯПТ и однородного простого языка чрезвычайно близки друг к другу.

§ 49. О составлении списка трансформаций применительно к конкретным языкам

Вспомним теперь, как мы интерпретировали в главе IV процесс формирования окрестностей. Точно так же можно, по-видимому, представить себе процесс составления списка трансформаций. Первоначально окрестности задавались лишь на основе единства семантической информации. Далее мы видели, что один из способов повышения степени однородности языка состоит в переопределении окрестностей. Мы разбили первоначально заданные окрестности так, чтобы они приближались к нашим подокрестностям (в идеальном случае совпадали с ними). Таким образом, окрестности были определены с учетом как семантической информации (первоначальное задание), так и системных отношений в языке (проба на однородность).

Применим данную процедуру к преобразованиям, которые мы хотим считать трансформациями.

Возьмем две фразы A_1 и B_1 , о которых мы, на основе анализа окрестностей, предполагаем, что они находятся в состоянии трансформируемости друг к другу. Очевидно, что не имеет смысла вводить трансформацию, применимую лишь к одной фразе данной структуры. Проверим поэтому, сколь близка предполагаемая трансформация к совершенной (в идеальном случае является ли она ей). Оставим только те трансформации, которые отвечают данному условию. Таким образом, на практике мы будем стремиться к тому, чтобы трансформация имела некоторый семантический инвариант и была достаточно близка к совершенной.

Осторожность в формулировке правил трансформации, требование приблизить практическое понятие трансформации к тому, которое мы имели в модели, важны потому, что при разумном использовании этих понятий можно разрешить целый

ряд лингвистически важных проблем, которые связаны с апеллированием к смыслу. В лингвистике и логике часто говорится о смысле высказывания без точного определения этого понятия. Наличие полного списка трансформаций для данного языка (или подязыка) позволяет точно определить, что значит, что фразы А и В имеют один и тот же смысл. Можно сказать, что А и В имеют один и тот же смысл, если существует цепочка фраз C_1, C_2, \dots, C_n , такая, что $C_1 = A, C_n = B$ и C_{i+1} получено из C_i путем применения некоторой трансформации.

Ясно, что установление формальных правил для идентификации смысла важно и для теории и для приложений.

В качестве иллюстрации того, какое значение имеет это обстоятельство для лингвистической теории, остановимся на вопросе синтаксических омонимов и синтаксических синонимов, т. е. именно тех вопросов, где рассматривается не только грамматическая форма, но и смысл.

О п р е д е л е н и я:

1) Некоторая конструкция А есть синтаксический омоним, если существуют две разные (по структуре дерева) фразы A_1 и A_2 , такие, что A_1 имеет тот же смысл, что и А, и A_2 имеет тот же смысл, что и А.

2) Две конструкции A_1 и A_2 суть синтаксические синонимы, если существует фраза А, имеющая тот же смысл, что A_1 , и тот же смысл, что A_2 , и при конфигурационном анализе обе конструкции A_1 и A_2 суть конфигурации с одним и тем же результирующим x .

П р и м е р ы. Конструкция *открытие лаборатории* есть синтаксический омоним, так как ее можно представить как возникшую в результате (близких к совершенным) трансформаций двух разных фраз: «*Лаборатория открыла (некто)*»; «*(Некто) открыл лабораторию*».

Конструкции «*(Он рассказал) о приезде брата*» и «*(Он рассказал), брат приехал*» суть синонимы, так как: 1) *приезд брата* → *брат приехал* есть (близкая к совершенной) трансформация; 2) обе конструкции имеют то же распределение, что и, например, слово *это*.

Эти примеры достаточно убедительно показывают значение трансформационного анализа (ТА), связанное с тем, что в основе ТА лежит идея семантической инвариантности. Они показывают, что ТА обещает решить кардинальную проблему языкознания, а именно проблему смысла, которая до сих пор не поддавалась решению в связи с неясностью путей формализации смысла. А там, где нет средств формализации, открывается простор для всяких субъективных построений. Только в наличии четких формальных правил установления тождества смысла — гарантия от субъективизма.

Между тем, следует откровенно сказать, что именно идея семантической инвариантности приводит к тому, что в ТА возникают существенные трудности, на некоторых из которых мы и остановимся в дальнейшем.

На первый взгляд преобразования, рассмотренные в ТА, мало отличаются от преобразований, рассмотренных при конфигурационном анализе, а именно: разрешается замена группы символов не одним результирующим, как в предыдущих моделях, а группой символов. Имеется, однако, существенное различие в том, как определяются конфигурации символов, подлежащие замене на данные символы. Если конфигурационный анализ целиком построен на чисто формальном критерии распределения, то для ТА анализ распределения может быть лишь вспомогательным средством.

Поскольку лингвист в ТА, как правило, лишен этого наиболее сильного формального критерия при составлении правил преобразования, то ему — если он откажется от выдвинутого в модели критерия совершенства — остается лишь одно: пользуясь собственной лингвистической интуицией, устанавливая, что в данном языке некоторая цепочка слов А трансформируется в некоторую другую цепочку В.

Правда, когда в фонологии или морфологии возникает ситуация, при которой необходимо выяснить, имеют ли два слова (группы слов) одно и то же значение, то лингвист, не смущаясь, обращается к носителю языка с вопросом: означают ли они одно и то же или нет. Без этого приема практически не обходится, в особенности в своей полевой работе, ни один лингвист.

Однако имеется существенная разница между тем, когда подобный прием применяется в фонологии и в ТА. Прежде всего обычно подчеркивается, что для достижения объективности вопрос, обращенный к носителю языка (informant), должен носить косвенный характер. Носитель языка должен «отвечать на вопрос, как что-либо говорится в его языке, но не о его языке»¹⁶. В частности, наиболее объективный результат дают вопросы типа: «правильно ли построена данная фраза?» или проще: «говорят ли так?». Ответ на вопрос о том, имеют ли два слова один и тот же смысл или нет, может уже давать весьма противоречивые результаты. Если, однако, такой вопрос задается при фонологических исследованиях, то возможная здесь субъективность не столь опасна, поскольку соответствующий критерий в фонологии может быть сделан дополнительным (или вообще отпасть, если фонема определяется чисто дистрибутивно).

¹⁶ Z. Harris. *Methods in structural linguistics*. Chicago, 1951, стр. 14.

Другое дело ТА. Здесь вопрос о том, значат ли фразы А и В одно и то же, есть вопрос не косвенный, а прямой, задаваемый прямо «в лоб». Поэтому опасность субъективного произвола достаточно велика. Эту трудность можно было бы преодолеть статистической обработкой данных опроса.

Но в ТА есть и другая трудность, которая подобным статистическим подходом не снимается. Дело в том, что в сложных структурах, составными элементами которых являются трансформированные фразы, появляются в результате соположения такие грамматические значения, которых не было в исходных фразах. Рассмотрим в качестве примера русские деепричастные обороты, например:

Так тяжкий млат, *дробя стекло*, кует булат

(Пушкин)

Если спросить носителя языка, означают ли слова *млат* (молот) *дробит стекло* и *дробя стекло* в данном случае одно и то же, то скорей всего мы получим отрицательный ответ. Более интеллигентный носитель языка, пожалуй, скажет, что в известной степени «млат *дробит стекло*» лишь часть значения оборота *дробя стекло*, что в обороте имеется уступительное значение и что оборот эквивалентен фразе «хотя он и *дробит стекло*». Такой ответ исследователя-лингвиста, несомненно, устраивает. Но можем ли мы быть уверены, что мы его всегда получим. Достаточно показать приведенную фразу десяти разным лицам, чтобы убедиться, насколько разноречива та информация, которую мы в подобных случаях получаем. При анализе подобных фраз наиболее точный результат лингвист получит, опираясь на собственную интуицию.

По-видимому, обе задачи — определение инвариантности смысла и составление правил трансформации — должны решаться одновременно. Составление правил должно фиксировать наши гипотезы, тем самым уточнять наши представления о семантических инвариантах данного языка, в то время как сами правила должны проверяться анализом внутрисистемных отношений. Именно здесь требование близости к совершенной трансформации чрезвычайно важно.

§ 50. Трансформационный синтез

Модель порождения, рассмотренная в § 46, состояла, как мы помним, в задании некоторых базисных структур и некоторых правил типа $x_i \rightarrow y_i$, где x_i всегда состоял из одного символа. Мы видели, что такая модель не дает возможности адекватно описать факты языка.

Пусть у нас имеются все конструкции языка в смысле § 44, т. е. S-структуры, в которых вычеркнуты все атрибуты. Отбе-

рем наиболее простые из этих конструкций, например соответствующие фразам:

- 1) *Мальчик идет,*
- 2) *Мальчик берет книгу,*
- 3) *Мальчик красивый.*

Множество подобных фраз простейшего строения Хомский назвал ядром языка. Можно считать, что остальные фразы получены из фраз ядра путем комбинаций способа развертывания фразы добавлением атрибутов и некоторых преобразований, которые Хомский назвал трансформациями, а мы будем пока называть «свободными преобразованиями». При свободном преобразовании некоторая цепочка символов x_1, \dots, x_m заменяется другой цепочкой y_1, \dots, y_n , причем число таких преобразований должно быть конечно.

Например, из фраз 1 и 2 можно получить такие:

1а) *Сегодня мальчик идет в кино* (путем замены *идет* на *идет сегодня* и *идет в кино* и трансформации *мальчик идет сегодня* → *сегодня мальчик идет*, где заменяются сразу три элемента);

2а) *Книга берется мальчиком* (путем пассивной трансформации, где также заменяются сразу три элемента).

При порождении правила свободного преобразования задаются произвольно, т. е. на правила замены одной цепочки другой не накладывается никаких ограничений.

Возникает, однако, вопрос, как целесообразней задавать правила свободного преобразования.

Прежде всего ясно, что эти правила невыгодно задавать для отдельных фраз. Мы ведь потребовали, чтобы число правил свободного преобразования было конечным, а хотим мы породить, разумеется, некоторое бесконечное множество фраз (ср. § 45).

Отсюда следует, что правила надо задавать для некоторых структур, причем, как ясно из рассуждений § 39, эти структуры должны быть совершенными.

Во-вторых, из приведенных в конце § 47 примеров следует, что мы хотим, чтобы наше преобразование сохраняло некоторый семантический инвариант.

Отсюда ясно, что свободное преобразование целесообразнее всего строить так, чтобы оно соответствовало трансформации из § 48; метод порождения свободными преобразованиями, обратный по отношению к трансформационному анализу, мы будем называть т р а н с ф о р м а ц и о н н ы м с и н т е з о м.

Здесь необходимо отметить, что мы в нашей работе несколько изменили изложение идей трансформационного метода по сравнению с работами Хомского. Трансформационный метод первоначально был сформулирован исходя из идеи порождения, т. е. изучались способы, как по некоторой фиксированной

фразе восстановить процесс ее порождения. Иначе говоря, трансформационный метод был задуман как распознающая модель в смысле § 4. Ясно, однако, что создание разумной распознающей модели предполагает наличие хороших моделей чистого анализа и чистого синтеза. Поэтому мы предпочли остановиться подробнее именно на аналитической модели, а трансформационный синтез изложить как простое обращение этой аналитической модели.

§ 51. К построению распознающей модели

В § 36 мы видели, что основные трудности, возникающие при построении распознающей модели, связаны с омонимией. Оказалось, что наиболее рациональным и наиболее общим способом разрешения омонимии является конфигурационный анализ.

Напомним, что при распознающей модели нам даны некоторые таблицы, позволяющие переходить от элементов текста к элементарным категориям. Как же быть в тех случаях, когда элементу текста соответствуют некоторая дизъюнкция элементарных категорий? Здесь предложены несколько моделей, которые, по-видимому, эквивалентны друг другу. Одна из таких моделей предложена Бар-Хиллелом¹⁷ на основе идей, высказанных польским логиком К. Айдукевичем¹⁸. Можно показать, что эта модель может быть переформулирована в терминах конфигурационного анализа¹⁹.

Мы рассмотрим сначала более простую модель²⁰. Возьмем следующее болгарское предложение:

Глаунавите мисли на весела вечер го различаха.

Здесь есть две омонимичные словоформы, т. е. словоформы, соответствующие дизъюнкциям элементарных категорий: *мисли* и *вечер*. Другие словоформы ставятся в соответствие одной элементарной категории каждая.

Имеем следующие соответствия между словами фразы и элементарными категориями:

глаунавите — прилаг., мн. ч., опред.

¹⁷ J. Bar-Hillel. A quasi-arithmetical notation for syntactic description. «Language», № 29, 1953, стр. 47—58.

¹⁸ K. Ajdukiewicz. Die syntaktische Konnexität. «Studia philosophica», Т. 1. Lwów, 1935.

¹⁹ См.: И. И. Ревзин. Установление синтаксических связей в МП методом Айдукевича — Бар-Хиллела и в терминах конфигурационного анализа. «Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста», вып. 2. М., 1961.

²⁰ Впервые эта модель была предложена автором в статье «Структуральная лингвистика, семантика и вопросы изучения слова». — ВЯ-1957, № 2, стр. 39—40.

мисли — (сущ., мн. ч., неопред.) \vee (гл., 3-е л., ед. ч., наст. вр.)

на — предлог

весела — прилаг., ед. ч., ж. р., неопред.

вечер — (сущ., ед. ч., ж. р., неопред.) \vee (наречие)

го — мест., ед. ч., вин. п., 3-е л., м. р.

различаха — гл., мн. ч., 3-е л., переходн., имперф.

Соединим теперь информации о каждом слове знаками конъюнкции. Получим следующую формулу:

(прилаг., мн. ч., опред.) & [(сущ., мн. ч., неопред.) \vee (гл., 3-е л., ед. ч., наст. вр.) &] (предлог) & (прилаг., ед. ч., ж. р., неопред.) & [(сущ., ед. ч., неопред.) \vee (наречие)] & (мест., ед. ч., вин. п., 3-е л., м. р.) & (гл., мн. ч., 3-е л., имперф.).

Если мы теперь установим, какие элементы этой формулы считать истинными, а какие ложными, то с формулой можно производить все операции, принятые в логике.

Мы, однако, поступим наоборот. Будем считать, что для операций дизъюнкции и конъюнкции над элементарными категориями имеют место два закона; ассоциативный и распределительный, и что введен ряд дополнительных формул, например:

1. (прилаг., мн. ч., опред.) & (сущ., мн. ч., неопред.) = сущ., мн. ч., опред.

2. (прилаг., ед. ч., ж. р., неопред.) & (сущ., ед. ч., ж. р., неопред.) = сущ., ед. ч., ж. р., неопред.

3. (сущ., мн. ч.) & (предлог) & (сущ.) = сущ., мн. ч.

4. (мест., вин. п.) & (гл., переходн.) = гл., непереходн.

Легко проверить, что применением указанных законов и преобразований 1—4 мы приходим к формуле:

5. (сущ., мн. ч., неопред.) & (гл., непереходн., имперф., 3-е л., мн. ч.).

Объявим формулу 5 одной из синтаксических аксиом болгарского языка (синтаксическая аксиома — совокупность элементарных категорий, соответствующих базисному упрощению). Сформулируем теперь следующее правило.

Если мы в результате сокращения по преобразованиям типа 1—4 пришли к синтаксической аксиоме, то все члены первоначальной формулы, входившие хотя бы в одно из преобразований, и все члены первоначальной формулы, входившие в синтаксическую аксиому, считаются истинными, а остальные считаются ложными.

Легко проверить, что применение этих правил позволяет разрешить омонимию слов *мисли* и *вечер*.

Ясно, что в основу этой модели положен конфигурационный анализ, а преобразования типа 1—4 соответствуют замене конфигураций результирующими.

Заметим, что понимание конфигурации как некоторой совокупности классов (типа выступающих в преобразованиях 1—4) было наиболее последовательно проведено Т. Н. Молошной в ее работах по теории машинного перевода²¹. Интересно, что Т. Н. Молошная называет конфигурациями и последовательности классов, соответствующие базисной структуре (типа 5). В самом деле, с точки зрения установления истинности элементарных категорий нет принципиальной разницы между формулами 1—4, с одной стороны, и формулой 5, с другой.

Для того, чтобы показать связь между изложенным методом и методом Айдукевича — Бар-Хиллела, изложим наш пример в терминах этих авторов.

Основные идеи работы К. Айдукевича сводятся к следующему²².

Вводятся две основные «семантические категории» (ср. § 23), категория имен (ей соответствует индекс n) и категория предложений (ей соответствует индекс s).

Слова, которые не принадлежат к этим двум основным категориям, рассматриваются как k -членные предикаты или функторы, и получают дробные индексы вида

$$\frac{s}{(\alpha_1) \dots (\alpha_m) [\beta_1] \dots [\beta_m]} \quad (1)$$

или

$$\frac{n}{(\alpha_1) \dots (\alpha_n) [\beta_1] \dots [\beta_r]} \quad (2)$$

где α_i, β_j — в свою очередь некоторые индексы, может быть и дробные²³. Смысл этих индексов в том, что предикат, имеющий индекс (1), или же функтор, имеющий индекс (2), образует с m словами, стоящими влево от него и имеющими индексы α_i , и с n словами, стоящими вправо от него и имеющими индексы β_j , предложение или, соответственно, имя.

Можно считать для общности, что в числителе не обязательно ставится n или s , а возможен любой индекс. Тогда правило принимает более простой вид. Выражению A приписывается индекс $\frac{\gamma}{(\alpha)[\beta]}$, если влево от A существует такое выражение B с индексом α и вправо от A — такое выражение C с индексом β , что выражению BAC приписывается индекс γ .

²¹ Т. Н. Молошная. Алгоритм перевода с английского языка на русский. «Проблемы кибернетики», вып. 3. М., 1960, стр. 209—272.

²² См.: K. Ajdukiewicz. Указ. соч., а также: R. Suszko. Zarys elementarnej skladni logicznej. Warszawa, 1957, стр. 26—27.

²³ J. Bar-Hillel. A quasi-arithmetical notation. «Language», N 29, 1953, стр. 47—58. Впоследствии Бар-Хиллел изменил способ записи, но мы здесь сохраним старую символику Айдукевича — Бар-Хиллела.

Распознающая модель строится как ряд предписаний типа

$$\alpha \frac{\gamma}{(\alpha)} \rightarrow \gamma \quad \text{или} \quad \frac{\gamma}{[\beta]} \beta \rightarrow \gamma.$$

Оставшаяся в результате всех сокращений цепочка называется экспонентом данного выражения. Экспонент, состоящий из одного первичного индекса n или s , называется правильным.

Выражение называется синтаксически связным, если его экспонент правилен.

Для простоты мы сначала рассмотрим случай, когда все конфигурации регулярны и каждая состоит ровно из двух слов, т. е. любое слово — или атрибут, или базисный элемент. Язык, в котором встречаются только такие слова, мы будем называть примитивным.

Предположим (в соответствии с требованием Айдукевича), что какой-то формальной процедурой выделены все собственные имена (например, для европейских языков все имена, которые пишутся с заглавной буквы) и каждому из них присвоен индекс n . Присвоим индекс n также всем словам, входящим в одно семейство с некоторым собственным именем.

Введем теперь следующие правила индексации:

1) Если одному из элементов базисной структуры некоторого предложения (в смысле определения § 44) присвоен индекс n , то второму элементу присвоим индекс $\frac{s}{[n]}$, если он стоял вправо от n и $\frac{s}{(n)}$, если он стоял влево от n .

2) Всякому атрибуту k элементу, имеющему индекс α , присваивается индекс $\frac{\alpha}{[\alpha]}$, если этот атрибут был правым, или $\frac{\alpha}{(\alpha)}$, если этот атрибут был левым.

Пусть теперь для каждого элемента языка получены применением указанной процедуры соответствующие индексы. Тогда каждой фразе соответствует некоторая последовательность индексов: i_1, i_2, \dots, i_n .

Распознающая модель строится следующим образом. Если в последовательности индексов стоят рядом два индекса $i_k i_{k+1}$, такие, что весь правый индекс совпадает со стоящим в круглых скобках знаменателем левого индекса или весь левый индекс совпадает со стоящим в квадратных скобках знаменателем правого индекса, то совпадающие индексы взаимно сокращаются (если знаменатель не снабжается скобками, то это означает, что положение индекса несущественно).

Например, пусть нам дана фраза

Очень маленькая девочка гуляла там.

Ей соответствует последовательность индексов

$$\frac{\frac{n}{(n)}}{\left(\frac{n}{n}\right)} \cdot \frac{n}{(n)} \cdot n \cdot \frac{s}{[n]} \cdot \frac{\frac{s}{n}}{\frac{s}{n}} \cdot \frac{s}{n}$$

Путем последовательных сокращений получаем

- 1) $\frac{n}{(n)} \cdot n \cdot \frac{s}{n}$,
- 2) $n \cdot \frac{s}{n}$,
- 3) s .

Аналогично для немецкого примера, приведенного Айдукевичем:

<i>Der</i>	<i>Flieder</i>	<i>duftet</i>	<i>sehr</i>	<i>stark</i>
$\frac{n}{(n)}$	n	$\frac{s}{n}$	$\frac{s}{n}$	$\frac{s}{n}$
			$\frac{s}{n}$	$\frac{s}{n}$
			n	n
			$\left[\begin{array}{c} \frac{s}{n} \\ \frac{n}{s} \\ \frac{s}{n} \end{array} \right]$	

получаем в конце концов s .

Посмотрим теперь, может ли данный метод быть распространен и на непримитивные языки. Мы еще не имеем для этого столь строгих правил как для примитивного языка, но чисто содержательно можно представить себе такую процедуру. Элементы, не являющиеся ни базисными, ни атрибутами, можно рассматривать как операторы, которые превращают слово определенного типа, т. е. имеющее определенный индекс, в слово другого типа. Так, предлоги превращают элемент

с индексом n в элемент с индексом $\frac{n}{s}$ или $\frac{n}{n}$, т. е. атрибут к

имени или глаголу; вспомогательные глаголы (типа болгарского *сѣм*) превращают имя или атрибут к имени в элемент с индексом $\frac{s}{n}$. Исходя из этого, предлогу может быть присвоен

индекс $\frac{s}{n}$ или $\frac{[n]}{(n)}$, вспомогательному глаголу — $\frac{s}{n}$
 $\frac{n}{(n)}$

или

$$\frac{s}{n}$$

$$\left[\begin{array}{c} \frac{s}{n} \\ \frac{s}{n} \\ \frac{s}{n} \end{array} \right]$$

и т. п. Так, например, фразе *Он поднялся на высокий холм* будет соответствовать следующая цепочка символов:

$$n \quad \frac{s}{n} \quad \frac{\frac{s}{n}}{(n)} \quad \frac{n}{(n)} \quad n.$$

Последовательным сокращением мы снова получаем *s*.

Вернемся теперь к рассмотренному выше болгарскому примеру. Пусть указанной выше процедурой слова получили следующие индексы:

$$\text{глупавите} \quad \frac{n}{(n)}$$

$$\text{мисли} \quad n, \frac{s}{n}$$

$$\text{на} \quad \frac{\frac{s}{n}}{(n)}, \quad \frac{n}{[n]}$$

$$\text{весела} \quad \frac{n}{(n)}$$

$$\text{вечер} \quad n, \frac{\frac{s}{n}}{s}$$

$$\text{го} \quad \frac{\frac{s}{n}}{(n) [n]}$$

$$\text{различаха} \quad \frac{s}{(n) [n]}$$

Интересно сравнить эту синтаксическую информацию с морфологической информацией, указанной выше. Отличие этого способа записи в том, что структура символов своей формой отражает вхождение в соответствующую конфигурацию.

Теперь сопоставим с каждым словом один из символов, приписываемых ему.

Получим восемь возможных последовательностей символов:

$$1) \frac{n}{(n)} n \frac{\frac{n}{[n]}}{\frac{n}{(n)}} n \frac{\frac{s}{n}}{\frac{s}{(n)}} \frac{s}{[n]},$$

$$2) \frac{n}{(n)} n \frac{\frac{\frac{s}{n}}{\frac{n}{(n)}}}{\frac{n}{(n)}} \frac{s}{n} \frac{\frac{s}{n}}{\frac{s}{(n)}} \frac{s}{(n)[n]},$$

$$3) \frac{n}{(n)} \frac{s}{n} \frac{\frac{\frac{s}{n}}{\frac{n}{(n)}}}{\frac{n}{(n)}} n \frac{\frac{s}{n}}{\frac{s}{(n)}} \frac{s}{(n)[n]},$$

$$4) \frac{n}{(n)} \frac{s}{n} \frac{\frac{\frac{s}{n}}{\frac{n}{(n)}}}{\frac{n}{(n)}} \frac{s}{n} \frac{\frac{s}{n}}{\frac{s}{(n)}} \frac{s}{(n)[n]} \text{ и т. д.}$$

Первая цепочка дает при сокращении

$$1a) n \frac{n}{[n]} \frac{s}{n}, \quad 1б) n \frac{s}{n}, \quad 1в) s.$$

Таким образом символы выбраны правильно.

Вторая цепочка дает

$$2a) n \frac{\frac{s}{n}}{\frac{n}{(n)}} \frac{n}{(n)} \frac{s}{n}$$

и дальше не сокращается. Значит, распознавание произведено неверно.

Третья цепочка дает

$$\frac{n}{(n)} \frac{s}{n} \frac{\frac{\frac{s}{n}}{\frac{n}{(n)}}}{\frac{n}{(n)}} \frac{n}{(n)} \frac{s}{n}$$

и далее опять не сокращается. Легко проверить, что и другие последовательности не приводятся к виду *s*. Мы получили таким образом ответ на поставленный вопрос.

В последнее время Бар-Хиллел, вслед за Хомским, всячески подчеркивает недостаточность его метода для адекватного описания реальных языков и в связи с этим ставит под сомнение возможность полностью автоматизованного высококачественного машинного перевода (*fully automatic high quality machine translation*, сокращенно *ФАНQMT*).

В частности, этот метод непригоден для всякого вида рачочных конструкций, когда два слова, относящихся друг к другу, существенно разделены какими-нибудь третьими словами.

В общем случае задача создания распознающей модели будет, видимо, решена, если удастся увязать ее с идеями трансформационного порождения. Это пока еще, сколько известно автору, никому не удалось.

§ 52. Модель порождения и парадигматическая модель

Мы рассмотрели целый ряд различных моделей языка. При этом читатель, по-видимому, уже заметил, что модели рассматриваются в порядке, который в определенном смысле согласуется с принятой в лингвистике традицией: фонология (парадигматика и синтагматика фонем), морфология (парадигматические отношения в грамматике) и синтаксис (синтагматические отношения в грамматике).

Важно выяснить теперь, отвечает ли этот порядок значимости тех или иных лингвистических явлений с точки зрения построения общей модели языкового поведения человека.

Прежде всего отметим, что в последнее время лингвисты и математики, занимающиеся формальным анализом языка, уделяют наибольшее внимание синтетическим, или порождающим, моделям, и это понятно: рассмотрение таких моделей явно (как в последних работах Хомского) или неявно связано с определенной гипотезой о процессе использования языка человеком. А именно этот процесс моделируется при помощи некоторого автомата того типа, какой разобран в § 47 в связи с гипотезой Ингве. В общем виде этот автомат может быть определен как система, способная порождать определенные последовательности символов. Наиболее общим типом такого автомата является устройство с неограниченной памятью, способное выполнить все потенциально осуществимые операции, необходимые для порождения любой заранее заданной линейной последовательности символов (такой автомат, по-видимому,

соответствует так называемой машине Тьюринга²⁴). Ясно, что человек, обладающий лишь ограниченной памятью, должен моделироваться автоматом, на деятельность которого наложен ряд ограничений. Наиболее сильные ограничения превращают такой автомат в устройство, порождающее язык с конечным числом состояний. Хомский показал, что между этими двумя крайними типами автоматов помещаются все автоматы, способные производить некоторое множество фраз.

Описанная точка зрения, несомненно, чрезвычайно плодотворна для развития языковедения и смежных дисциплин, которые в последнее время начали широко пользоваться кибернетическими представлениями (например, психология).

Как мы видели, конструирование порождающих, т. е. синтетических, моделей естественно связать с рассмотрением аналитических моделей, причем каждая синтетическая модель может быть построена на основе некоторой аналитической модели, помогающей осмыслить основные понятия и отношения и дающей объективные критерии для составления тех или иных правил преобразования (например, правил развертывания конфигураций или же трансформационных правил). При построении аналитических моделей может быть существенно использован опыт дескриптивной лингвистики, для многих понятий которой (например, «ядро», «спутник», «непосредственно составляющий» и т. п.) можно построить довольно естественный аналог в терминах аналитической, а затем и синтетической модели.

Тот факт, что понятия, отработанные в процессе развития одного из направлений в лингвистике, могут быть использованы в некоторой модели, несомненно, повышает познавательную ценность соответствующей модели.

Констатируя, однако, плодотворность представления о языке, связанного с конструированием автомата описанного выше вида, следует все же заметить, что если свести все различия между разными моделями к каким-то ограничениям на характер выполняемых автоматом операций, то мы вряд ли приблизимся к созданию модели, более или менее адекватно отражающей языковое поведение человека. Показательно, что при подобном подходе к моделированию языка не остается места для всего того, что было накоплено новейшим европейским языковедением, начиная с работ Бодуэна де Куртена и де Соссюра через пражский лингвистический кружок и вплоть до современных школ европейского структурализма.

Основное достижение этого направления, по-видимому, в понимании языка как «системы» — концепции, в разработку которой значительный вклад внесли и советские языковеды.

²⁴ См. описание «машины Тьюринга» в кн.: Б. А. Трахтенброт. Алгоритмы и машинное решение задач, 2-е изд. М., 1960, стр. 62—100.

Представители этого направления — каковы бы ни были расхождения между ними — считают, что отдельные элементы языка вступают в сложную иерархию парадигматических отношений, причем лингвистическая значимость отдельного элемента определяется не только его местом в некоторой речевой последовательности, но и общим характером системы, наличием или отсутствием в ней других единиц и теми отношениями внутри системы, пересечением которых является данная единица.

Эти соображения заставляют внимательнее проанализировать описанную выше гипотезу порождения языка автоматом.

В этой гипотезе до сих пор, по-видимому, неявно предполагалось, что происходит одно из двух:

а) автомат состоит лишь из устройства порождения и сам по собственной программе производит фразы;

б) автомат перерабатывает фразы (т. е. некоторые последовательности сигналов), получаемые извне.

В моделях порождения, о которых идет здесь речь, рассматривается лишь первый случай. И это вполне закономерно, если исходить из того, что во втором случае фразы, перерабатываемые автоматом, производятся некоторым другим автоматом того же вида, причем информация идет только в одном направлении: от автомата, производящего фразы, к автомату, перерабатывающему их, и не может создаваться перерабатывающим автоматом. Ибо в этом случае соединение двух автоматов одного типа приводит к модели, которая изоморфна модели одного автомата, произвольно производящего некоторые фразы. Это можно выразить и несколько иначе. Гипотеза, согласно которой производство человеческой речи может моделироваться автоматом описанного выше типа, предполагает пассивную переработку информации, при которой может происходить лишь потеря информации, а отнюдь не обогащение ее.

Собственно говоря, у нас нет никаких прямых доказательств того, что функционирование языка должно мыслиться иначе. Однако вполне можно представить себе другой подход к моделированию, при котором исследуются не только ограничения, связанные с характером операций (замены одной цепочки символов другой), но и формулируются некоторые гипотезы о внутренней структуре автомата, связанные с процессом активной переработки информации, повышения степени ее организованности.

Эта более сложная модель может быть представлена как состоящая из двух частей: автомата, перерабатывающего информацию, и источника информации, но в отличие от рассмотренной выше ситуации мы не делаем никаких гипотез относительно источника информации, зато мы предполагаем, что перерабатывающий автомат воспринимает любую информацию как искаженную под влиянием шума, может быть частично утерянную, или вообще имеющую неудобный для переработки вид.

В связи с этим перерабатывающий автомат распадается на две части: устройство порождения и устройство упорядочения информации. Схематически это можно представить себе следующим образом:



Мы будем считать, что упорядочивающее устройство получает информацию от источника, распределяет ее на основании каких-то критериев по классам (например, категориям, окрестностям) и передает ее порождающему устройству, которое в свою очередь передает в упорядочивающее устройство информацию о тех элементах, которые заняли предшествующие места в производимых отрезках фразы. Эта дополнительная информация дает возможность окончательной категоризации (например, выбор семейства в наших моделях).

Можно было бы считать также, что упорядочивающее устройство связано обратной связью и с источником информации, влияя уже на характер производимой этим источником цепочки символов. Такое допущение даже необходимо в том случае, если автомат не только активно воздействует на получаемую информацию, но и одновременно изменяет окружающую среду, как это происходит в процессе человеческого труда. Но с точки зрения моделирования это, по-видимому, не вносит существенно нового и поэтому мы ограничимся лишь гипотезой активной переработки информации некоторым упорядочивающим устройством (сделанное здесь замечание, однако, проливает некоторый свет на процесс вторичного упорядочения, о котором несколько ниже).

Устройство, упорядочивающее и обогащающее информацию, группирующее ее по определенным категориям, принадлежность к которым, вообще говоря, не задана ее первоначальной формой, соответствовало бы тогда функционированию парадигматической системы языка.

При этом можно представить себе, что необходимость борьбы против энтропических тенденций²⁵ приводит к тому, что сама система разных классов, служащих для упорядочения

²⁵ См.: Н. Винер. Кибернетика и общество. Перевод с англ. М., 1958, стр. 45—62.

(в терминах наших моделей «семейства», «окрестности», «категории»), в свою очередь подвергается строгому упорядочению (такое, например, стремление к «однородности языка»).

Необходимо отметить, что подобное вторичное упорядочение, вообще говоря, не является необходимым для передачи информации, поскольку оно резко увеличивает избыточность (в теоретико-информационном смысле), но оно может быть, вероятно, объяснено из общего стремления к преодолению энтропических тенденций окружающего мира. Ибо именно это стремление характерно для всей человеческой деятельности. Как мы знаем, основу этой деятельности составляет труд, сыгравший решающую роль в возникновении языка. Итак, вовсе не необходимо, чтобы язык функционировал по модели, снабженной упорядочивающим устройством. Имеется, однако, много косвенных соображений в пользу того, что эта модель ближе к реальному функционированию языка, чем та, в которой информация перерабатывается пассивно и поэтому исследуются лишь ограничения на вид операций. В самом деле, факты совершенно различной категоризации отрезков действительности в разных языках свидетельствуют о том, что в сущности ни одна из категорий не является абсолютно необходимой. Так каждая из рассмотренных нами категорий имени в русском языке отсутствует в каком-нибудь другом языке: род отсутствует в английском, падеж и число в индонезийском и т. д. Тем не менее нет языков, в которых отсутствовали бы вообще категории. Более того нет языков, где категории не образовывали бы определенной стройной системы, достаточно близкой к «однородной». Объяснить наличие категорий только требованиями ограничения ранга (как мы пытались это сделать в связи с гипотезой Ингве) вряд ли удастся в общем случае.

Что касается вторичного упорядочения внутри самой системы категорий, то отмеченный нами факт изоморфизма (или во всяком случае далеко идущего сходства) между фонологией и грамматикой подтверждает, что стремление к системности не является чем-то случайным или внешним.

Можно было бы предположить, что система упорядочения информации наделена такими же гомеостатическими свойствами (т. е. стремлением к сохранению внутренней устойчивости²⁶), как и другие системы, служащие для сопротивления организма силам, направленным к повышению общей энтропии окружающей среды.

Во всяком случае мы видели, что, исходя из подобной предпосылки (которая, правда, как мы уже сказали, не является необходимой), можно моделировать процесс формирования

²⁶ По поводу понятия «гомеостат» см.: У. Р о с с Э ш б и. Введение в кибернетику. Перевод с англ. М., 1959, стр. 123.

индивидуальных парадигм как в ходе обучения языку, так и в историческом развитии некоторого языка.

Для лингвиста основное преимущество модели, снабженной устройством упорядочения информации, в том, что в подобной модели могут быть широко использованы идеи и представления, развиваемые современным языковедением, кардинальным понятием которого, несомненно, является понятие «системы» (примером может служить совершенно естественно включение представлений Мартине в парадигматическую модель).

Кроме того, рассмотрение такой модели поможет, по-видимому, моделировать процесс перевода как процесс перехода от одной системы упорядочения к другой²⁷.

Наконец, можно надеяться, что эта модель даст возможность построения — на кибернетических основаниях — строгой типологии языков, которая будет основана на последовательном сравнении упорядочивающих устройств. По-видимому, может быть создан универсальный набор признаков, исходя из которых можно получить самую разнообразную категоризацию, учитывающую особенности всех языков мира. Во всяком случае начатое в настоящее время изучение дифференциальных признаков на разных уровнях языка, несомненно, поможет значительно продвинуться вперед в исследовании различных классов упорядочивающих устройств.

Представляется, что модель, включающая устройство упорядочения, более соответствует и описанию языка с точки зрения смежных дисциплин, и в первую очередь психологии.

Если первая модель в какой-то мере отражает атомистический подход американского бихейвиоризма, с которым, несомненно, связана и дескриптивная лингвистика, то вторая модель скорее соответствует современным устремлениям к изучению «целостей», начатому еще так называемой гештальт-психологией, которая в известной мере близка европейскому структурализму 30-х годов.

Во всяком случае было бы интересно попытаться начать строить модели порождения с вработанными в них устройствами упорядочения. Между тем увлечение «чистыми» моделями порождения привело к тому, что парадигматические модели разработаны далеко не столь строго и последовательно как модели порождения. Больше того, теоретико-множественные концепции, наиболее удобные для парадигматического анализа языка, не развиваются должным образом.

Хотя нам и не удалось связать модель порождения с парадигматическими (первые попытки в этом направлении чита-

²⁷ См. И. И. Ревзин, В. Ю. Розенцвейг. К обоснованию лингвистической теории перевода. — ВЯ, 1962, № 1.

тель может усмотреть в разделе о трансформационном анализе и трансформационном синтезе), мы все же остановились на таком порядке изложения теории моделей языка, который соответствует необходимости исследования парадигматических моделей перед тем, как строятся синтагматические модели и, в частности, модели порождения.

Традиционная иерархия уровней языка отражает интуитивно верно угаданное значение отдельных частей друг для друга. А именно, синтаксис мыслим и без морфологии, но это обеднит его возможности. Точно так же можно построить морфологию, не прибегая ни к синтагматике, ни к парадигматике фонем. Однако все развитие современного языковедения показывает, что хорошая морфология немислима без фонологии, а хороший синтаксис немислим без морфологии.

Предложенная схема моделирования языкового поведения, разумеется, является односторонней, поскольку она недостаточно учитывает коммуникативную функцию языка (которая должна моделироваться по крайней мере двумя автоматами, передающими информацию друг другу). Однако для этого необходимо широкое внедрение теоретико-информационных представлений в моделирование языка, а это принципиально выходит за рамки нашей работы, в которой рассматриваются лишь логические (множественные), но не статистические модели языка.

Примечание. Ю. Д. Апресян указал автору на то, что для лингвистической интерпретации трансформационной модели из § 48 целесообразно ввести дополнительное условие трансформируемости фразы А во фразу В. Он предложил потребовать, чтобы кроме условий 1 и 2 из § 48 выполнялось следующее: для любых слов x_1 и x_2 таких, что $S(x_1)$ и $S(x_2)$ суть составляющие в $S(A)$ и что имеет место отношение $S(x_1) \Rightarrow S(x_2)$ в смысле § 43, должны найтись в В слова y_1 , входящее в $\Gamma(x_1)$ и y_2 , входящее в $\Gamma(x_2)$ такие, что $S(y_1) \Rightarrow S(y_2)$, если $S(x_1)$ и $S(x_2)$ не есть предикативная пара, и $S(y_1) \Rightarrow S(y_2)$ или $S(y_2) \Rightarrow S(y_1)$ — в противном случае. Можно удостовериться, что этим не затрагиваются пары фраз, которые мы хотели бы считать трансформируемыми друг в друга, но зато исключаются такие пары фраз, как *брат жены приехал* и *жена брата приехала* или *Павел сидит в гостях у Ивана* и *Иван сидит в гостях у Павла*.

ПРИЛОЖЕНИЕ

К главе I, § 3

Пример гетеронимии

Рассмотрим язык, в котором имеется три элемента: a , b и c . Отмеченными считаются все кортежи четной длины, например ab , ba , $baab$, $babaca$ и т. п.

Порождение такого множества отмеченных кортежей можно представить двумя способами:

1) из элементов a , b и c образуется произвольным способом некоторое выражение длины $n = 1, 2, 3, \dots$, затем подсчитывается длина кортежа — если соответствующее число четное, то кортеж отмечен, если нет, то прибавляется некоторый элемент. Получившийся кортеж считается отмеченным;

2) из элементов a , b , c образуются последовательности длины 2, а именно: aa , ab , ba , bb , bc , cb , cc , ac , ca .

Любая комбинация этих последовательностей длины $n = 1, 2, 3 \dots$ (n — число последовательностей) есть отмеченный кортеж.

К главе II, § 13

Теорема 2.1. Если подсистема не содержит пустых клеток, то она фонологически однородна.

Доказательство: Пусть $z \in K^{(1)}(x) \cap K^{(2)}(y)$. Это означает, что z и x имеют общую архифонему α и могут быть представлены в виде

$$x = \alpha \cup m_1, \quad (1)$$

$$z = \alpha \cup m_j. \quad (2)$$

Аналогично y и z имеют общую архифонему β ($\alpha \neq \beta$) и могут быть представлены в виде

$$y = \beta \cup l_1, \quad (3)$$

$$z = \beta \cup l_j. \quad (4)$$

Из (2) и (4) следует, что $l_j \in \alpha$ и $m_j \in \beta$. Иначе говоря, $\alpha = \alpha' \cup l_j$ и $\beta = \beta' \cup m_j$.

$$z = \alpha' \cup l_j \cup m_j = \beta' \cup m_j \cup l_j \quad (5)$$

Обозначим $\alpha' = \beta' = \gamma$ и равенства (1) — (4) перепишем в виде

$$x = \gamma \cup I_1 \cup m_1, \quad (1a)$$

$$z = \gamma \cup I_2 \cup m_2, \quad (2a)$$

$$y = \gamma \cup I_1 \cup m_2. \quad (3a)$$

Сформируем фонему $w = \gamma \cup I_1 \cup m_1$ (по условию I_1 и I_2 полностью однородны). Поскольку при разбиении $K^{(1)}$ x и z попадают в один класс, то y и w также попадут в один класс $K^{(1)}(y) = K^{(1)}(w)$, поскольку они имеют общую архифонему $\gamma \cup I_1$ и, кроме того, w и x имеют общую архифонему $\gamma \cup m_1$, а z и y имеют общую архифонему $\gamma \cup m_2$. Итак, $w \in K^{(1)}(y)$. Аналогично доказывается, что $w \in K^{(2)}(x)$. Отсюда следует, что $K^{(1)}(y) \cap K^{(2)}(x)$ непусто. Теорема доказана.

Теорема 2.2. Если подсистема фонологически однородна, то два регулярных класса одного разбиения не могут содержать разного количества фонем.

Доказательство. Пусть при данном разбиении два класса $K^{(1)}(x)$ и $K^{(1)}(y)$ содержат неодинаковое число фонем:

$$K^{(1)}(x) = \{x_1, x_2, \dots, x_n\},$$

$$K^{(1)}(y) = \{y_1, y_2, \dots, y_m\},$$

и пусть для определенности $n > m$. Заметим, что если мы докажем наше утверждение для случая, когда некоторый член из $K^{(1)}(x)$, например x_1 , имеет общую архифонему с членом из $K^{(1)}(y)$, например y_1 , то оно автоматически распространяется на общий случай, поскольку любые два класса данного разбиения (по построению) можно связать цепочкой классов, каждая пара из которых удовлетворяет указанному свойству. Итак, рассмотрим наш специальный случай. При некотором разбиении $K^{(2)}$ x_1 и y_1 входят в один класс. При этом разбиении, поскольку их классов больше, а x_i и x_j не могут входить вместе в $K^{(2)}(y)$, то найдется x_i , такой, что не существует $y_1 \in K^{(2)}(x_i)$. Пусть для определенности таким x будет x_n .

Тогда

$$K^{(2)}(x_n) \cap K^{(1)}(y_1) = \emptyset.$$

В то же время

$$K^{(1)}(x_n) \cap K^{(2)}(y_1) = x_1.$$

Отсюда следует, что язык не является фонологически однородным. Теорема доказана.

Теорема 2.3. Если система фонологически однородна и x составляет изолированный класс при разбиении $K^{(1)}$ и $y \in K^{(0)}(x)$, где $K^{(0)}$ — исходное разбиение, то $K^{(1)}(y)$ есть изолированный класс.

Доказательство. Пусть $K^{(1)}(x)$ — изолированный класс. По условию существует y , такой, что $y \in K^{(0)}(x)$. Покажем, что $K^{(1)}(y)$ есть изолированный класс. Допустим противное. Пусть $z \in K^{(1)}(y)$. Тогда $y \in (K^{(1)}(z) \cap K^{(0)}(x))$ и должна найтись (в силу фонологической однородности) фонема w , такая, что $w \in (K^{(1)}(x) \cap K^{(0)}(z))$. Отсюда, поскольку $K^{(1)}(x)$ — изолированный класс, $w = x$ и $z \in K^{(0)}(x)$. Отсюда $z \in K^{(0)}(y)$ и z и y имеют две разные общие архифонемы: α , определяющую разбиение $K^{(0)}$, и β , определяющую разбиение $K^{(1)}$, а по лемме 1 (стр. 172) это невозможно. Теорема доказана.

Теорема 2.4. В фонологически однородной подсистеме две соседние фоне́мы имеют одинаковый ранг.

Доказательство. Пусть x_1 и y_1 две соседние фонемы, пусть $y_1, x_2, x_3, \dots, x_{m-1}$ (I) все соседние с x_1 фонемы, $x_1, y_2, y_3, \dots, y_{n-1}$ (II) все соседние с y_1 фонемы и $n > m$.

Мы будем говорить, что y_j ($2 \leq j \leq n-1$) находится в отношении корреляции с x_i ($2 \leq i \leq m-1$) по паре (x_1, y_1) , если для каких-то разных разбиений k и l $x_1 \in K^{(k)}(y_1) \cap K^{(l)}(x_1)$, а $y_j \in K^{(k)}(x_1) \cap K^{(l)}(y_1)$.

Если мы докажем, что для каждого x_i найдется одно и только одно y_j (и обратно: для каждого y_j одно и только одно x_i), удовлетворяющее этим условиям, то отсюда следует, что можно установить взаимно-однозначное соответствие между (I) и (II), а значит $m = n$ и теорема будет доказана.

Возьмем теперь x_i ($2 \leq i \leq m-1$); x_i и x_1 имеют общую архифонему α , определяющую некоторое разбиение $K^{(l)}$, и $x_1 \in K^{(l)}(x_i)$.

Кроме того, x_1 имеет общую архифонему с y_1 . Если эта архифонема есть α , то x_i является соседней фонемой с y_1 , т. е. x_i равно одному из y_j . Такие элементы мы просто поставим в соответствие самим себе. Пусть теперь x_1 и y_1 имеют общую архифонему β ($\beta \neq \alpha$), определяющую разбиение $K^{(k)}$, т. е. $x_1 \in K^{(k)}(y_1) \cap K^{(l)}(x_1)$. По условию однородности должна существовать фонема $z \in K^{(k)}(x_1) \cap K^{(l)}(y_1)$. Раз $z \in K^{(l)}(y_1)$, то z — соседняя с y_1 фонема. Если $z = y_1$, то $x_1 \in K^{(k)}(x_1)$, а это противоречит лемме 1. Если $z = x_1$, то $z \in K^{(k)}(y_1) \cap K^{(l)}(y_1)$, но этого не может быть, так как мы предположили, что α и β разные архифонемы. Итак $z = y_j$ ($2 \leq j \leq n-1$), и y_j находится в отношении корреляции к x_i по паре (x_1, y_1) . Пусть теперь две разные фонемы z' и z'' находятся в отношении корреляции к x_i по паре (x_1, y_1) . Тогда имеют место следующие соотношения:

$$z' \in K^{(k)}(x_1) \cap K^{(l)}(y_1),$$

$$z'' \in K^{(k)}(x_1) \cap K^{(l)}(y_1).$$

Отсюда следует, что z' и z'' принадлежат к одному классу при разбиении $K^{(k)}$ и к одному классу при разбиении $K^{(1)}$, а этого не может быть в силу леммы 1 (стр. 172).

Аналогично доказывается, что для каждой фонемы y_j найдется только одна фонема x_i , находящаяся с ней в отношении корреляции по паре (x_i, y_i) .

Таким образом каждой фонеме из (I) ставится в соответствие одна фонема из (II) и обратно. Теорема доказана.

Теорема 2.5. Если подсистема является полной и фонологически однородной, то она не содержит пустых клеток.

Доказательство. Пусть выполнены условия теоремы и пусть имеются признаки m_i и m_j , такие, что для некоторой архифонемы α и $\alpha \cup m_i$ и $\alpha \cup m_j$ суть фонемы. Кроме того, пусть имеется фонема $\beta \cup m_i$, но отсутствует фонема $\beta \cup m_j$ (на месте этой фонемы пустая клетка). Пусть $\alpha = a \cup \lambda$, $\beta = b \cup \lambda$, т. е. предположим сначала, что мы имеем три фонемы, попарно различающиеся только одним признаком:

$$F_1 = a \cup \lambda \cup m_i,$$

$$F_2 = a \cup \lambda \cup m_j,$$

$$F_3 = b \cup \lambda \cup m_i.$$

F_1 и F_2 входят в один класс при разбиении, определяемом архифонемой α . Пусть это будет $K^{(1)}(F_1) = K^{(1)}(F_2)$. F_1 и F_3 входят в один класс при разбиении, определяемом архифонемой $m_i \cup \lambda$. Пусть это будет $K^{(2)}(F_1) = K^{(2)}(F_3)$. По условию однородности должна существовать фонема F_4 , входящая в $K^{(1)}(F_3)$ и в $K^{(2)}(F_2)$. Пусть эта фонема имеет вид

$$F_4 = x \cup \lambda \cup y.$$

F_4 должна иметь общую архифонему с F_2 . Отсюда для одного из элементов x или y , пусть для x , верно, что $x = a$ или $x = m_j$. Но если $x = a$, то F_4 имеет общую архифонему с F_1 , что невозможно. Итак,

$$F_4 = m_j \cup \lambda \cup y.$$

Поскольку F_4 должна иметь общую архифонему с F_3 , то $y = m_i$ или $y = b$. Первое невозможно, так как однородные признаки несовместимы. Итак,

$$F_4 = m_j \cup \lambda \cup b = \beta \cup m_j,$$

и предположение об отсутствии $\beta \cup m_j$ опровергнуто. Пусть теперь фонемы F_1 и F_3 различаются k признаками и пусть для фонем, различающихся $k-1$ признаками, утверждение теоремы доказано.

Тогда существуют три фонемы:

$$F_1 = m_i \cup w \cup a_1 \cup a_2 \cup \dots \cup a_k,$$

$$F_2 = m_j \cup w \cup a_1 \cup a_2 \cup \dots \cup a_k,$$

$$F_3 = m_i \cup w \cup b_1 \cup b_2 \cup \dots \cup b_k.$$

Предположим, что не существует фонемы

$$F_4 = m_j \cup w \cup b_1 \cup b_2 \cup \dots \cup b_k.$$

По условию полноты существует фонема F_3' , такая, что F_3 отличается от F_3' одним признаком, а от F_1 $k-1$ признаками.

По индуктивному предположению из существования F_1 , F_2 и F_3' вида

$$F_3' = m_i \cup w \cup a_1 \cup b_2 \cup b_3 \cup \dots \cup b_k$$

следует существование F_4' вида

$$F_4' = m_j \cup w \cup a_1 \cup b_2 \cup b_3 \cup \dots \cup b_k.$$

А если F_3' , F_4' и F_3 попарно различаются только одним признаком, то, как мы доказали, должна существовать фонема F_4 вида:

$$F_4 = m_j \cup w \cup b_1 \cup b_2 \cup \dots \cup b_k.$$

Теорема доказана.

Л е м м а 1. Пусть $x \in K^{(i)}(y)$, тогда для любого j ($j \neq i$) $x \in K^{(j)}(y)$.

Пусть $x \in K^{(i)}(y)$ и $x \in K^{(j)}(y)$. Возьмем первый класс j -го разбиения $K_1^{(j)}$. Пару фонем (x, y) можно связать цепочкой пар с парой фонем (z, w) , входящих в $K_1^{(j)}$, причем для соседних пар цепочки верно, что внутри пары две фонемы имеют общую архифонему и что для каждой фонемы одной пары найдется фонема в другой паре, имеющая с ней общую архифонему. Индукцией по длине этой цепочки доказывается, что $z \in K^{(i)}(w)$, а это противоречит пункту 4 нашего построения.

С л е д с т в и е. Любая архифонема однозначно определяет некоторое разбиение.

К главе IV, § 27

Теорема 4.1 (Кулагина). Для любого x верно: $V^*(x) = V^{**}(x)$.

Прежде чем доказывать эту теорему, введем несколько вспомогательных понятий (которые понадобятся нам и в дальнейшем).

Рассмотрим два разбиения множества Ξ , а именно $V^{(1)}$ и $V^{(2)}$.

Мы будем говорить, что разбиение $V^{(1)}$ крупнее, чем разбиение $V^{(2)}$, или $V^{(1)}$ является укрупнением разбиения $V^{(2)}$, если $V^{(1)}(x) \supseteq V^{(2)}(x)$ для всех $x \in \Xi$, причем $V^{(1)}$ и $V^{(2)}$ не тождественны.

Мы будем говорить, что разбиение $V^{(1)}$ есть регулярное укрупнение $V^{(2)}$, если из условия $V^{(2)}(y) \subseteq V^{(1)}(x)$ следует, что $V^{(2)}(y) \underset{V^{(2)}}{\sim} V^{(2)}(x)$,

иначе говоря, если каждое множество разбиения $V^{(1)}$ есть сумма $V^{(2)}$ -эквивалентных между собой элементов разбиения $V^{(2)}$. Частным случаем регулярного укрупнения и является, например, производное разбиение. Прежде всего докажем ряд вспомогательных утверждений.

Лемма 1. Пусть $B^{(1)}$ есть регулярное укрупнение $B^{(2)}$ и пусть какая-то $B^{(1)}$ -структура соответствует какой-то $B^{(2)}$ -структуре (соответствует в том смысле, что $B_i^{(2)} \subseteq B_i^{(1)}$, где i — номер места элемента данного разбиения в структуре). Тогда отмеченной $B^{(1)}$ -структуре не может соответствовать неотмеченная $B^{(2)}$ -структура.

Доказательство. Пусть отмеченная $B^{(1)}$ -структура имеет вид

$$B^{(1)}(x_1) \ B^{(1)}(x_2) \ \dots \ B^{(1)}(x_n), \quad (1)$$

где $A = x_1 x_2 \dots x_n$ была той отмеченной фразой, которая и привела к отмеченности структуры (1).

Возьмем $B^{(2)}$ -структуру этой фразы:

$$B^{(2)}(x_1) \ B^{(2)}(x_2) \ \dots \ B^{(2)}(x_n). \quad (2)$$

Она, разумеется, отмечена. Но в лемме предполагается, что существует хотя бы одна такая $B^{(2)}$ -структура:

$$B^{(2)}(u_1) \ B^{(2)}(u_2) \ \dots \ B^{(2)}(u_n), \quad (3)$$

которая, с одной стороны, не отмечена и, с другой стороны, состоит из $B_i^{(2)}$, таких, что $B_i^{(2)} \subseteq B_i^{(1)}$.

Поскольку $B^{(1)}$ есть регулярное укрупнение $B^{(2)}$, то $B^{(2)}(x_1) \sim_{B^{(2)}} B^{(2)}(u_1)$

Это означает, что, заменив в (3) $B^{(2)}(u_1)$ на $B^{(2)}(x_1)$, мы вновь получили бы неотмеченную структуру. Проведя n раз такую замену, мы вновь должны получить неотмеченную структуру. На самом же деле такой заменой мы перейдем от (3) к (2). Итак, (2) является одновременно отмеченной и неотмеченной структурой.

Полученное противоречие доказывает лемму.

Лемма 2. Пусть $B^{(1)}$ — регулярное укрупнение $B^{(2)}$. Тогда из $B^{(2)}(x) \sim_{B^{(2)}} B^{(2)}(y)$ следует $B^{(1)}(x) \sim_{B^{(1)}} B^{(1)}(y)$ и из $B^{(1)}(x) \sim_{B^{(1)}} B^{(1)}(y)$ следует $B^{(2)}(u) \sim_{B^{(2)}} B^{(2)}(w)$ для любых u и w , таких, что $B^{(2)}(u) \subseteq B^{(1)}(x)$ и $B^{(2)}(w) \subseteq B^{(1)}(y)$.

Доказательство. Докажем первое утверждение. Предположим, что $B^{(1)}(x)$ не $B^{(1)}$ -эквивалентно $B^{(1)}(y)$ в условиях леммы, т. е. имеются две такие фразы

$$A = x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_n$$

и

$$A' = x_1 x_2 \dots x_{i-1} y x_{i+1} \dots x_n,$$

что из $B^{(1)}$ -структур $B^{(1)}(A)$ и $B^{(1)}(A')$ одна отмечена, а другая нет, в то время как $B^{(2)}(A)$ и $B^{(2)}(A')$ одновременно отмечены или не отмечены. Пусть $B^{(1)}(A)$ отмечена, тогда по лемме 1 отмечена и $B^{(2)}(A)$. Отсюда

по условию отмечена и $V^{(2)}(A')$. По лемме 1 $V^{(1)}(A')$ также должна быть отмеченной структурой, а мы предположили, что она не отмечена. Тем самым мы пришли к противоречию. Первая часть утверждения доказана.

Докажем вторую часть утверждения. Пусть отмечена $V^{(2)}$ -структура:

$$V^{(2)}(A_1) \quad V^{(2)}(u) \quad V^{(2)}(A_2), \quad (1)$$

где $V^{(2)}(u) \subseteq V^{(1)}(x)$.

Из отмеченности (1) следует, что отмечена и $V^{(1)}$ -структура:

$$V^{(1)}(A_1) \quad V^{(1)}(x) \quad V^{(1)}(A_2). \quad (2)$$

Тогда по условию леммы отмечена и $V^{(1)}$ -структура:

$$V^{(1)}(A_1) \quad V^{(1)}(y) \quad V^{(1)}(A_2). \quad (3)$$

Возьмем теперь $V^{(2)}$ -структуру:

$$V^{(2)}(A_1) \quad V^{(2)}(w) \quad V^{(2)}(A_2), \quad (4)$$

где $V^{(2)}(w) \subseteq V^{(1)}(x)$. Поскольку $V^{(1)}$ есть регулярное укрупнение $V^{(2)}$, то (4) обязательно отмечена по лемме 1.

Таким образом, мы видим, что из отмеченности (1) всегда следует отмеченность (4). Аналогичным путем показывается, что из отмеченности (4) для произвольных u и w следует отмеченность (1). Тем самым лемма доказана.

Лемма 3. Пусть $V^{(1)}$ — регулярное укрупнение $V^{(2)}$. Тогда $V^{(1)*}(x) = V^{(2)*}(x)$ для любого x (иначе говоря, при регулярном укрупнении производное разбиение по $V^{(1)}$ совпадает с производным разбиением по $V^{(2)}$).

Доказательство. Возьмем $x \in V^{(2)*}(y)$. Докажем, что $x \in V^{(1)*}(y)$. По лемме 2 $V^{(1)}(x) \underset{V^{(1)}}{\sim} V^{(1)}(y)$, т. е. $x \in V^{(1)}(y)$. Итак, мы доказали, что

$$V^{(2)*}(x) \subseteq V^{(1)*}(x). \quad (1)$$

Пусть теперь $x \in V^{(1)*}(y)$. Докажем, что $x \in V^{(2)*}(y)$. Итак, $V^{(1)}(x) \underset{V^{(1)}}{\sim} V^{(1)}(y)$. По лемме 2 отсюда следует, что $V^{(2)}(u) \underset{V^{(2)}}{\sim} V^{(2)}(w)$ для любых u и w , таких, что $V^{(2)}(u) \subseteq V^{(1)}(x)$ и $V^{(2)}(w) \subseteq V^{(1)}(y)$, т. е. в частности $V^{(2)}(x) \underset{V^{(2)}}{\sim} V^{(2)}(y)$. Отсюда $x \in V^{(2)}(y)$, т. е.

$$V^{(1)*}(x) \subseteq V^{(2)*}(x). \quad (2)$$

Включения (1) и (2) полностью доказывают лемму 3.

Доказательство теоремы 4.1. Мы знаем, что производное разбиение есть частный случай регулярного укрупнения. Отсюда очевидно, что $V^{**}(x) = V^*(x)$ для любого x .

Пример искусственного языка, в котором два элемента принадлежат к одному семейству, но разным типам:

Пусть нам задан язык, состоящий из пяти элементов:

$$x, y_1, y_2, z_1, z_2.$$

Пусть нам заданы девять отмеченных фраз:

$$\begin{array}{ll} A_1 = y_2x & A_6 = z_1x \\ A_2 = y_2y_1 & A_7 = z_1z_2 \\ A_3 = y_1y_2 & A_8 = xy_2 \\ A_4 = xz_1 & A_9 = z_1y_1 \\ A_5 = y_1z_1 & \end{array}$$

и следующее разбиение на окрестности:

$$\begin{array}{l} \Gamma(x) = x, \\ \Gamma(y) = y_1, y_2, \\ \Gamma(z) = z_1, z_2. \end{array}$$

Сопоставление A_1 и A_2 , A_3 и A_8 , A_4 и A_5 , A_6 и A_9 приводит нас к заключению, что

$$y_1 \sim x \text{ или } x \in S(y_1).$$

Сопоставим с этими фразами их Γ -структуры. Имеем следующие отмеченные Γ -структуры:

$$\begin{array}{ll} \Gamma(A_1) = \Gamma(y) \Gamma(x) & \Gamma(A_6) = \Gamma(z) \Gamma(x) \\ \Gamma(A_3) = \Gamma(A_2) = \Gamma(y) \Gamma(y) & \Gamma(A_7) = \Gamma(z) \Gamma(z) \\ \Gamma(A_4) = \Gamma(x) \Gamma(z) & \Gamma(A_8) = \Gamma(x) \Gamma(y) \\ \Gamma(A_5) = \Gamma(y) \Gamma(z) & \Gamma(A_9) = \Gamma(z) \Gamma(y) \end{array}$$

Сопоставляя

$$\begin{array}{ll} \Gamma(A_1) \text{ и } \Gamma(A_6) & \Gamma(A_7) \text{ и } \Gamma(A_5) \\ \Gamma(A_2) \text{ и } \Gamma(A_5) & \Gamma(A_8) \text{ и } \Gamma(A_4) \\ \Gamma(A_2) \text{ и } \Gamma(A_7) & \Gamma(A_9) \text{ и } \Gamma(A_2) \end{array}$$

мы приходим к выводу, что $\Gamma(z) \underset{\Gamma}{\sim} \Gamma(y)$.

С другой стороны, сопоставляя отмеченную структуру $\Gamma(y) \Gamma(x)$ с неотмеченной $\Gamma(x) \Gamma(x)$, мы приходим к выводу, что $\Gamma(y)$ не Γ -эквивалентно $\Gamma(x)$.

Теорема 4.2. Для правильных языков для любого x :

- 1) $R(x) \subseteq T(x)$,
- 2) $T(x) = R^*(x)$.

Доказательство.

1) Пусть x и y принадлежат к одному и тому же разделу. Это значит, что между ними существовала цепочка $\Gamma_1, \Gamma_2, \dots, \Gamma_n$, такая, что

при всяком i элементы Γ_i и Γ_{i+1} принадлежали либо одной и той же окрестности, либо одному и тому же семейству. Если Γ_i и Γ_{i+1} принадлежали одной окрестности, то они должны принадлежать одному и тому же типу, так как разбиение на типы есть производное от разбиения на окрестности. Если же Γ_i и Γ_{i+1} принадлежали одному семейству, то они принадлежат к одному типу из-за правильности. Таким образом, все элементы цепочки остаются внутри типа; первое утверждение доказано.

2) Прежде чем доказывать вторую часть теоремы, докажем следующее вспомогательное утверждение:

Л е м м а 4. Разбиение на разделы в правильном языке совпадает с разбиением на окрестности или является регулярным укрупнением разбиения на окрестности.

Ясно, что раздел может просто совпасть с окрестностью. (Это верно и для неправильного языка.) Пусть теперь раздел крупнее окрестности. Тогда в раздел входит несколько окрестностей $\Gamma_1, \Gamma_2, \dots, \Gamma_n$. По теореме 4.2 (пункт 1) весь раздел содержится в одном типе, а это и означает, что $\Gamma_i \sim \Gamma_k$. Отсюда разбиение на разделы есть (для случая, когда раздел крупнее типа) регулярное укрупнение разбиения на окрестности.

С л е д с т в и е. Применяя лемму 3, мы (для случая, когда раздел не совпадает с окрестностью) получаем

$$R^*(x) = \Gamma^*(x) = T(x).$$

Это же равенство тривиальным образом выполняется, если раздел совпадает с окрестностью. Теорема доказана. \square

Теперь покажем на разобранным нами примере, что требование правильности языка существенно. Легко видеть, что в рассмотренном нами примере получаются следующие типы:

$$T_1 = x,$$

$$T_2 = y_1, y_2, z_1, z_2,$$

и следующие разделы:

$$R_1 = x, y_1, y_2,$$

$$R_2 = z_1, z_2.$$

Сопоставим теперь каждую из фраз A_i , например A_1, A_4, A_6, A_7 , с соответствующей R -структурой. Получим, что все возможные R -структуры из двух элементов, например

$$R(A_1) = R_1R_1$$

$$R(A_4) = R_1R_2$$

$$R(A_6) = R_2R_1,$$

$$R(A_7) = R_2R_2,$$

отмечены. Отсюда $R_1 \widetilde{R} R_2$ и все пять элементов входят в $R^*(x)$, т. е. $T(x) \subseteq \subseteq R^*(x)$ и $T(z) \subseteq R^*(z)$. С другой стороны, например, $x \notin T(z)$, т. е. здесь производное от разбиения на разделы не совпадает с разбиением на типы.

Главе IV, § 28

Теорема 4.3 (Успенский). Всякий формально однородный язык есть правильный язык.

Доказательство. Пусть $x \in S(y)$ и пусть для любого элемента $x_1 \in \Gamma(x)$ верно, что из непустоты пересечения $\Gamma(x_1) \cap S(y) = x$ следует наличие элемента z , такого, что $S(x_1) \cap \Gamma(y) = z$, т. е. для любых a_1 и b_1 из отмеченности

$$a_1 a_2 \dots a_k x_1 a_{k+1} \dots a_n \quad (1)$$

следует отмеченность

$$a_1 a_2 \dots a_k z a_{k+1} \dots a_n \quad (2)$$

и из отмеченности

$$b_1 \dots b_1 z b_{1+1} \dots b_m \quad (3)$$

следует отмеченность

$$b_1 \dots b_1 x_1 b_{1+1} \dots b_m. \quad (4)$$

Взяв соответствующие Γ -структуры, мы сразу же убедимся, что из отмеченности $\Gamma(a_1) \dots \Gamma(a_k) \Gamma(x) \Gamma(a_{k+1}) \dots \Gamma(a_n)$ следует отмеченность $\Gamma(a_1) \dots \Gamma(a_k) \Gamma(y) \Gamma(a_{k+1}) \dots \Gamma(a_n)$ и обратно. Отсюда $\Gamma(x) \underset{\Gamma}{\sim} \Gamma(y)$, что и требовалось доказать.

Примечание. Условие однородности вовсе не является необходимым для того, чтобы $\Gamma(x) \underset{\Gamma}{\sim} \Gamma(y)$. Достаточно, чтобы для любого $x_1 \in \Gamma(x)$ нашелся элемент $z \in \Gamma(y)$ и такие элементы $a'_1 \in \Gamma(a_1)$, что из отмеченности $a_1 \dots a_k x_1 a_{k+1} \dots a_n$ следует отмеченность $a'_1 \dots a'_k z a'_{k+1} \dots a'_n$ и обратное верно для любого элемента $y' \in \Gamma(y)$. Например, легко проверить, что $\Gamma(\text{окно}) \underset{\Gamma}{\sim} \Gamma(\text{метро})$, хотя ни одно из слов, входящих в $\Gamma(\text{окно})$, не эквивалентно слову *метро*.

Главе IV, § 29

Теорема 4.4: а) морфологические подклассы образуют разбиение всего множества слов; б) производное разбиение от разбиения на морфологические подклассы совпадает с разбиением на типы.

Доказательство. а) Обозначим морфологический подкласс, соответствующий y , через $M(y)$. Пусть $x \in M(y)$, т. е. существует w , такое, что $w \in \text{ПIS}(x) \cap \Gamma(y)$. По определению подсемейства существует z такое, что $z \in S(y) \cap \Gamma(x)$. Пусть $z \in \text{ПIS}(y)$, тогда найдется слово $a \in \Gamma(y)$, для которого не существует d , такого, что $d \in S(a) \cap \Gamma(z)$, но $\Gamma(y) = \Gamma(w)$, а $\Gamma(z) = \Gamma(x)$. Итак, $w \in \Gamma(a) \cap S(x)$, но нет b , такого, что $b \in S(a) \cap \Gamma(x)$. Это, однако, противоречит предположению, что $w \in \text{ПIS}(x)$. Итак, $z \in \text{ПIS}(y)$. Отсюда следует, что если $x \in M(y)$, то $y \in M(x)$.

Пусть теперь $x \in M(y)$ и $y \in M(z)$. Иначе говоря, существуют a и b , такие, что $a \in \text{ПIS}(x) \cap \Gamma(y)$ и $b \in \text{ПIS}(y) \cap \Gamma(z)$. Таким образом, $y \in \Gamma(a) \cap \text{ПIS}(b)$.

По доказанному выше отсюда следует существование w , такого, что $w \in \text{PS}(a) \cap \Gamma(b)$. Но $\text{PS}(a) = \text{PS}(x)$, а $\Gamma(b) = \Gamma(z)$. Итак, $x \in M(z)$. Первая часть утверждения доказана.

б) Докажем сначала, что M -разбиение есть регулярное укрупнение Γ -разбиения. Пусть $\Gamma(x) \subseteq M(y)$. Покажем, что $\Gamma(x) \not\sim \Gamma(y)$, т. е. что для произвольных фраз $A = a_1 \dots a_k$ и $B = b_1 \dots b_m$ структуры $\Gamma(A) \Gamma(x) \Gamma(B)$ и $\Gamma(A) \Gamma(y) \Gamma(B)$ одновременно отмечены или одновременно не отмечены. Пусть первая отмечена, а вторая нет.

Не ограничивая общности, можно считать, что фраза $a_1 \dots a_k x b_1 \dots b_m$ отмечена. Тогда из условия $x \in M(y)$ следует существование w , такого, что $w \in \text{PS}(x) \cap \Gamma(y)$. Раз $w \in \text{S}(x)$, то фраза $a_1 \dots a_k w b_1 \dots b_m$ отмечена. Γ -структура этой фразы есть $\Gamma(A) \Gamma(w) \Gamma(B)$, и она отмечена как структура отмеченной фразы. Но поскольку $\Gamma(w) = \Gamma(y)$, эта структура по предположению не отмечена. Полученное противоречие показывает, что $\Gamma(x) \not\sim \Gamma(y)$. Применив теперь лемму 3, мы получаем, что $M^*(x) = T(x)$.

Теорема 4.5. Если язык регулярен, то два слова, относящиеся к одному морфологическому подклассу, входят в одни и те же парадигматические категории.

Доказательство. Пусть $x \in M(y)$ и $K(x)$ — произвольная парадигматическая категория, содержащая x . По условию существует слово w , такое, что $w \in \text{PS}(x) \cap \Gamma(y)$. В силу регулярности $w \in K(x)$, а из $w \in \Gamma(y)$ и определения парадигматической категории следует, что $y \in K(w)$, где $K(w)$ — произвольная парадигматическая категория, содержащая w . Итак, $y \in K(x)$. Теорема доказана.

Теорема 4.6. Если язык формально однороден и регулярен, то в нем все непарадигматические категории распределены равномерно.

Доказательство. Пусть имеется непарадигматическая категория K , такая, что $x \in N(y)$, $z \in K \cap \Gamma(x)$ и $K \cap \Gamma(y) = 0$. Раз $x \in N(y)$ и по предположению $\Gamma(x) = \Gamma(z)$, то $\Gamma(z) \cap \text{S}(y) \neq 0$. Отсюда в силу однородности языка $\text{S}(z) \cap \Gamma(y) \neq 0$ (1). В силу формальной однородности семейство совпадает с подсемейством. В силу регулярности $\text{S}(z) \subseteq K$. Из (1) следует, что имеется w , такое, что $w \in \text{S}(z) \cap \Gamma(y)$, и, значит, $w \in K \cap \Gamma(y)$, а это противоречит нашему предположению. Теорема доказана.

Глава IV, § 35

Теорема 4.7. В элементарном языке из условия $S_1 \rightarrow S_j$ следует, что хотя бы одна S_j -фраза не является S_1 -фразой.

Доказательство. Предположим противное. Пусть при $S_1 \rightarrow S_j$ всякая S_j -фраза является S_1 -фразой. Пусть $x \in S_1$ и $y \in S_j$. Поскольку $\text{S}(x) \neq \text{S}(y)$, то имеется хотя бы одна отмеченная S_j -фраза $A_1 y A_2$, такая, что соответствующая S_1 -фраза $A_1 x A_2$ не отмечена.

$S_1 \rightarrow S_j$, и по нашему предположению S_j -фраза $A_1 y A_2$ одновременно является S_1 -фразой, т. е. замена какого-то элемента на x должна дать отмеченную фразу.

Поскольку $A_1 x A_2$ не отмечена, то должна получиться отмеченная фраза, содержащая как y , так и x , т. е. два слова из одной элементарной категории. Это противоречит элементарности языка. Тем самым утверждение теоремы доказано.

Главе V, § 38

Теорема 5.1. Язык, в который входит хотя бы одна фраза с рамкой (i, j) , не является ограниченным языком с конечным числом состояний.

Доказательство. Пусть $A = x_1 \dots x_i x_{i+1} x_{i+2} \dots x_{j-2} x_{j-1} x_j \dots x_n$ есть отмеченная фраза с рамкой (i, j) . Тогда фраза $A = x_1 \dots y_i x_{i+1} \dots \dots x_{j-1} y_j x_{j+1} \dots x_n$ должна быть также отмечена. Между x_i и x_j имеется по крайней мере два слова x_k и x_{k+1} ($i < k < k+1 < j$), стоящие рядом на одинаковом расстоянии от начала в обеих фразах. Возьмем $k = j - 2$. В силу наложенного на язык ограничения при производстве обеих фраз мы пройдем через состояние $C_{k, k+1}$. Но при переходе в следующее состояние из этого должны быть две возможности перехода в другое состояние, а именно при переходе в одно из них должно производиться x_j , а при переходе в другое y_j .

Но тогда фраза $A = x_1 \dots x_i x_{i+1} \dots x_{j-1} y_j \dots x_n$ также будет произведена нашей схемой, т. е. должна быть отмеченной, а это противоречит условию. Теорема доказана.

Примечание. Метод, использованный при доказательстве, можно распространить и на случай, когда на язык не наложено никаких ограничений; а именно, можно показать, что если имеется хотя бы одна фраза с рамкой длины m , то нужно 2^m различных состояний для производства всех отмеченных и только отмеченных фраз данного языка. Это и делается Хомским.

Главе V, § 39

Теорема 5.2: а) Всякая отмеченная S-структура есть совершенная B-структура; б) Если при некотором фиксированном разбиении B всякая B-структура или не отмечена или совершенна, то для любого элемента x верно $B(x) \subseteq S(x)$.

Доказательство.

а) Пусть дана некоторая отмеченная S-структура $S(x_1) S(x_2) \dots S(x_n)$. Без ограничения общности можно считать, что именно фраза $x_1 x_2 \dots x_n$ отмечена. Докажем, что данная S-структура совершенна, т. е. произвольная фраза $y_1 y_2 \dots y_n$ отмечена, если только $x_1 \in S(y_1)$. Пусть $y_1 y_2 \dots y_n$ не отмечена. Заменим $y_2 y_3 \dots y_n$ на $x_2 x_3 \dots x_n$, тогда получившаяся фраза $y_1 x_2 \dots x_n$ не отмечена в силу $x_2 \in S(y_2), \dots, x_n \in S(y_n)$ и отмечена в силу $x_1 \in S(y_1)$. Полученное противоречие доказывает наше утверждение.

б) Пусть задано некоторое разбиение B, при котором произвольная B-структура совершенна или не отмечена, и пусть $x \in B(y)$. Докажем,

что $x \in S(y)$. Возьмем произвольную фразу

$$Cx D. \quad (1)$$

Если она отмечена, то отмечена и В-структура

$$B(C) B(x) B(D), \quad (2)$$

а значит, в силу совершенности (2) и фразы

$$Cy D. \quad (3)$$

Аналогично, если не отмечена (1), то не отмечены (2) и (3). Итак, (1) и (3) одновременно отмечены или не отмечены, а значит $x \in S(y)$.

Главе V, § 40

Теорема 5.3. Если S_i и S_j — суть результирующие конфигурации k -го ранга $\tilde{S}_{(k)}$, то $S_{i(k)} S_j$.

Доказательство. Возьмем произвольную отмеченную S -структуру $(k-1)$ -го ранга, в которой встретилось S_j , пусть она имеет вид

$$S(A) S_j S(B). \quad (1)$$

Тогда отмечена и

$$S(A) \tilde{S}_{(k)} S(B). \quad (2)$$

Но тогда отмечена и

$$S(A) S_i S(B). \quad (3)$$

Обратно, если отмечена (3), то отмечена (2) и (1). Отсюда $S_{i(k)} S_j$, что и требовалось доказать.

Главе V, § 41

Теорема 5.4. Для того, чтобы S -структура имела однозначно определенную норму, достаточно, чтобы в процессе сокращения ранг сокращенных конфигураций не превышал ранга вновь возникающих конфигураций. Чтобы доказать, что в условиях теоремы и при нашем способе сокращения все базисные упрощения данной S -структуры состоят из одного и того же количества элементов, достаточно доказать следующее вспомогательное утверждение:

Лемма 1. Пусть дана S -структура $(k-1)$ -го ранга: $A\tilde{S}_{(k)}B$, где $\tilde{S}_{(k)}$ — конфигурация k -го ранга с результирующими $S_k^1, S_k^2, \dots, S_k^n$ и пусть S_k^i образует с группой элементов Z' из A и группой элементов Z'' из B конфигурацию m -го ранга $S_{(m)} = Z' S_k^i Z''$ (где Z' или Z'' , но, разумеется, не то и другое, может представлять собой пустую S -структуру с результирующими $S_m^1, S_m^2, \dots, S_m^1$, и пусть $m \geq k$). Тогда $S_k^j (1 \leq j \leq n$ и $i \neq j)$ также образует с Z' и Z'' конфигурацию m -го ранга $S'_{(m)} = Z' S_k^j Z''$ с теми же результирующими $S_m^1, S_m^2, \dots, S_m^1$.

Доказательство. По определению конфигурации структуры $(m-1)$ -го ранга

$$CZ' S_k^1 Z''D \quad (1)$$

и

$$CS_m^1 D \quad (2)$$

одновременно отмечены или не отмечены для произвольных структур m -го ранга C и D . Раз $m \geq k$, структуры m -го ранга CZ' и $Z''D$ одновременно являются и структурами k -го ранга. Между тем по теореме 5.3 $S_{k(k)}^1 \sim S_k^1$. Отсюда структуры k -го ранга

$$CZ' S_k^1 Z''D \quad (3)$$

и (1) одновременно отмечены или не отмечены для произвольных C и D k -го ранга (в том числе и для случая $CZ' = A$ и $Z''D = B$). Отсюда структуры (2) и (3) одновременно отмечены или не отмечены при произвольных C и D , а значит элемент S_k^1 входит в конфигурацию S_m^1 с теми же результирующими, что у конфигурации S_m^1 . Тем самым лемма доказана.

Заметим, что для случая, когда конфигурации выделяются как справа налево, так и слева направо, доказательство теоремы не проходит. Легко построить такой пример:

Пусть $S_1 S_2 S_3 S_4 S_5$ есть заданная S -структура, и $S_1 S_2 \rightarrow S'_k, S_2 S_3 \rightarrow S''_k, S'_k S_4 S_5 \rightarrow \bar{S}$ и других конфигураций нет. Тогда исходная S -структура имеет две нормы: $N_1 = 2$ (количество элементов в упрощении $S_1 \bar{S}$) и $N_2 = 4$ (количество элементов в упрощении $S'_k S_3 S_4 S_5$). Для реальных языков такого примера построить нельзя, поэтому введенное нами ограничение вполне естественно.

Теорема 5.5. Если в языке множество рангов отмеченных базисных упрощений ограничено, все базисные упрощения отмеченных S -структур простые и какая-то отмеченная S -структура имеет норму 1, то и все отмеченные S -структуры данного языка имеют норму 1.

Доказательство (от противного). Возьмем какую-нибудь S -структуру нормы k , ее базисное упрощение имеет вид $S_1 S_2 \dots S_k$. Пусть базисное упрощение S -структуры нормы 1 имеет вид \bar{S} . По условию имеется некоторый максимальный ранг. В силу нашего предположения $S_1 S_2 \dots S_k$ не может быть конфигурацией с результирующим \bar{S} , т. е. имеются такие A и B , что из структур g -го ранга

$$S(A) S_1 S_2 \dots S_k S(B) \quad (1)$$

$$S(A) \bar{S} S(B) \quad (2)$$

одна отмечена, а другая не отмечена. Пусть для определенности отмечена (1). Заметим, что по нашему выбору ранга g и (1) и (2) суть базисные

упрощения. Но поскольку (1) и (2) — простые структуры, то $S(A)$ и $S(B)$ должны быть пустыми. Но тогда отмечена и (2). Мы же предполагали, что структура (2) не отмечена. Полученное противоречие доказывает теорему.

Главе V, § 42

Теорема 5.6. Существует способ сокращения любой S -структуры до базисного упрощения, при котором выделяются только регулярные конфигурации и предбазисные конфигурации.

Прежде чем доказывать теорему, введем следующие понятия. Мы будем считать, что конфигурация \tilde{S}_i непосредственно порождает конфигурацию \tilde{S}_j , если один из результирующих конфигурации \tilde{S}_i входит в качестве составной части в конфигурацию \tilde{S}_j . Мы будем говорить, что конфигурация \tilde{S}_1 порождает конфигурацию \tilde{S}_j , если имеется цепочка конфигураций: z_1, z_2, \dots, z_n , такая, что $z_1 = S_i$, $z_n = S_j$ и z_k непосредственно порождает z_{k+1} для любого k ($1 \leq k \leq n$).

Докажем теперь следующие два свойства.

Лемма 2. Если $\tilde{S}^{(1)}$ — конфигурация k -го ранга с результирующим S_α непосредственно порождает конфигурацию m -го ранга $\tilde{S}^{(2)}$ с результирующим S_β , то S -структура, получающаяся из $S^{(2)}$ путем замены S_α на $\tilde{S}^{(1)}$, представляет собой конфигурацию n -го ранга с результирующим S_β , где $n = m$, если $m \geq k$, и $n = k$, если $m < k$.

Доказательство. Рассмотрим S -структуры

$$S(A_1) S(B_1) \tilde{S}^{(1)} S(B_2) S(A_2), \quad (1)$$

$$S(A_1) S(B_1) S_\alpha S(B_2) S(A_2), \quad (2)$$

$$S(A_1) S_\beta S(A_2), \quad (3)$$

где $S(B_1) S_\alpha S(B_2) = \tilde{S}^{(2)}$ (причем $S(B_1)$ или $S(B_2)$, но не обе, могут быть пустыми S -структурами), а $S(A_1)$, $S(A_2)$, $S(B_1)$, $S(B_2)$ — структуры $(n-1)$ -го ранга. Очевидно, что (1) и (2) одновременно отмечены при произвольных A_1, A_2, B_1 и B_2 при нашем выборе n . Отсюда следует, что (1) и (3) одновременно отмечены или не отмечены при любых $S(A_1)$ и $S(A_2)$ $(n-1)$ -го ранга, т. е. $S(B_1) \tilde{S}^{(1)} S(B_2)$ есть конфигурация $(n-1)$ -го ранга, что и требовалось доказать.

Лемма 3. Всякая нерегулярная конфигурация порождает какую-нибудь регулярную конфигурацию или предбазисную конфигурацию.

Доказательство. Возьмем некоторую нерегулярную конфигурацию, результирующий ее входит в некоторую другую, и т. д., так как этот процесс в силу конечности исходной S -структуры не может продолжаться бесконечно, то мы обязательно дойдем до порождения некоторой предбазисной конфигурации. Тем самым лемма доказана (ясно,

что раньше чем дойти до предбазисной конфигурации, мы могли породить просто регулярную конфигурацию).

Теперь укажем способ сокращения, при котором будут получаться только регулярные конфигурации или предбазисные конфигурации, те и другие мы будем называть нормальными. Выделим некоторую минимальную конфигурацию S_1 . Если она нормальна, то сокращаем ее. Если она ненормальна, объединим ее с порождаемой ей конфигурацией \bar{S}_2 способом, указанным в лемме 2, и т. д., пока мы не дойдем до некоторой нормальной, которую сокращаем, и т. д. Ясно, что все конфигурации при таком сокращении будут нормальными. Теорема, таким образом, доказана.

К главе V, § 43

Теорема 5.7. Для любых двух составляющих R_1 и R_2 данного предложения имеет место одно из двух: 1) один из них непосредственно подчиняет другой; 2) имеется составляющий R_3 , который подчиняет R_2 и одновременно подчиняет R_1 .

Доказательство. Любой составляющий является или базисным составляющим (собственным или несобственным) или атрибутивным составляющим. Если оба составляющих базисные, то для них по определению предложения выполняется первое условие.

Пусть теперь один из составляющих, например R_1 , не является базисным. Тогда он является атрибутивным составляющим и непосредственно подчинен некоторому другому; тот в свою очередь является атрибутивным или базисным, и т. д. В конце концов мы дойдем до базисного составляющего R_{β_1} , такого, что R_1 подчинен R_{β_1} . Если $R_2 = R_{\beta_1}$, то первое условие выполнено. Пусть $R_2 \neq R_{\beta_1}$. Если R_2 — базисный составляющий, то R_{β_1} подчиняет R_2 и в то же время R_{β_1} подчиняет R_1 , т. е. второе условие выполнено. Пусть наконец R_2 не есть базисный составляющий. Повторив рассуждение, проведенное в начале для R_1 , мы увидим, что имеется базисный составляющий R_{β_2} , такой, что R_2 подчиняет R_{β_2} . R_{β_2} или совпадает с R_{β_1} , или подчиняет R_{β_1} . В обоих случаях R_{β_2} удовлетворяет второму условию. Теорема доказана.

Теорема 5.8. В предложении имеется одна и только одна предикативная пара.

Доказательство. Рассмотрим процесс сокращения S-структуры, являющейся предложением. Если бы ее базисное упрощение состояло только из одного элемента S_1 , то ему соответствовал бы некоторый составляющий R_1 (само S_1 или некоторая предбазисная конфигурация с результирующим S_1), для которого отсутствует другой базисный составляющий и, стало быть, нет составляющего R_2 , такого, что R_2 подчиняет R_1 , а это противоречит определению предложения. Пусть теперь базисное упрощение состоит из n элементов, где $n > 2$. Тогда им соответствует n базисных составляющих, и каждый из них подчинен ровно $(n - 1)$ составляющим, что противоречит условию. Остается только одна возможность, а именно, что базисное упрощение состоит ровно из двух элементов и, стало быть, имеется ровно два базисных составляющих. Теорема доказана.

Теорема 5.9. Если отмеченная S-структура имеет норму 2, то она является предложением.

Доказательство (от противного). Пусть S-структура имеет норму 2, и для какого-то составляющего R_1 не найдется составляющего R_2 , такого, что $R_2 \rightarrow R_1$. Тогда R_1 не может быть атрибутивным составляющим. Пусть теперь R_1 есть базисный составляющий. Но тогда он, по нашему предположению, единственный базисный составляющий, и S-структура имеет норму 1.

К главе V, § 45

Теорема 5.10. Если язык задан как конечное множество отмеченных фраз, то в нем не может быть выделено ни одной регулярной конфигурации.

Доказательство (от противного). Пусть в нашем языке имеется хотя бы одна регулярная конфигурация, для простоты возьмем конфигурацию из двух элементов V_1V_2 с результирующим V_2 . Это означает, что для любых структур фраз A_1 и A_2 одновременно отмечены V-структуры

$$V(A_1) V_1V_2V(A_2) \quad (1)$$

и

$$V(A_1) V_2V(A_2), \quad (2)$$

т. е., иначе говоря, где бы ни встретился элемент V_2 , всегда можно поставить перед ним элемент V_1 так, что получившаяся структура будет отмечена. Возьмем структуру (1) и заменим в ней элемент V_2 на V_1V_2 , получившаяся структура

$$V(A_1) V_1V_1 V_2V(A_2) \quad (3)$$

также должна быть отмеченной. Вместо одного элемента V_1 у нас имеется V_1V_1 . Повторим ту же операцию со структурой (3). Так как множество фраз конечно, в конце концов мы дойдем до такой V-структуры

$$V(A_1) \underbrace{V_1 \dots V_1}_{k \text{ раз}} V_2V(A_2), \quad (4)$$

которая отмечена, в то время как V-структура

$$V(A_1) \underbrace{V_1 \dots V_1}_{k+1 \text{ раз}} V_2V(A_2) \quad (5)$$

уже не может быть отмеченной.

Обозначим $V(A_1) \underbrace{V_1 \dots V_1}_{k \text{ раз}}$ как $V(C)$. Тогда (4) и (5) примут вид:

$$V(C) V_2V(A_2), \quad (4a)$$

$$V(C) V_1V_2V(A_2). \quad (5a)$$

При этом (4a) есть структура отмеченная, а (5a) неотмеченная. Это противоречит предположению о том, что V_1V_2 есть конфигурация с результирующим V_2 . Теорема доказана.

Главе V, § 46

Теорема 5.11 (Хомский). Всякий язык с конечным числом состояний есть терминальный язык, но не всякий терминальный язык есть язык с конечным числом состояний.

Доказательство. Пусть имеется n состояний C_1, C_2, \dots, C_n , и пусть некоторым парам состояний C_i, C_j соответствуют слова a_{ijx} . Тогда грамматику терминального языка можно построить следующим образом:

$$\begin{aligned} C_i &\rightarrow a_{ijx} C_j \quad (i, j = 0, 1, \dots, n); \\ C_0 &\rightarrow \text{пустое слово.} \end{aligned}$$

Число этих правил определяется числом пар связанных состояний, которое конечно.

Теперь приведем пример языка, который является терминальным языком, но не является языком с конечным числом состояний. Пусть в этот язык входят все фразы вида $ab, aabb, aaabbb$ и вообще $\underbrace{a \dots a}_{n \text{ раз}}$

$\underbrace{bb \dots b}_{n \text{ раз}}$ и только они.

Все эти фразы можно получить как терминальный язык из двух правил вывода:

$$\begin{aligned} Z &\rightarrow ab, \\ Z &\rightarrow aZb. \end{aligned}$$

Между тем, этот язык не является языком с конечным числом состояний. В самом деле, пусть устройство имеет N состояний. Тогда фраза длины, превосходящей N , не может быть построена без петли, т. е. возвращения в некоторое состояние, которое уже было пройдено. Между тем, любой цикл в правой части приведет к производству фраз вида $\underbrace{a \dots a}_{n \text{ раз}} \underbrace{b \dots b}_{n+k}$, которые не отмечены. Аналогично для левой части. Цикл, включающий ab , приведет к неотмеченным фразам вида $aa \dots aab \dots abbb \dots b$. Итак, не существует устройства с конечным числом состояний, производящего только фразы нашего языка. Теорема доказана.

Главе V, § 48

Теорема 5.12. Если в простом и однородном языке некоторое преобразование удовлетворяет требованию семантической связи, то оно есть трансформация.

Прежде всего докажем следующее вспомогательное утверждение:

Лемма 4. Если слово x в простом однородном языке входит в единичную окрестность, то любое слово y , такое, что $y \in S(x)$, также входит в единичную окрестность.

Доказательство. Пусть $\Gamma(x) = \{x\}$ и $y \in S(x)$. Пусть $y' \neq y$ и $y' \in \Gamma(y)$. Отсюда следует, что $y \in \Gamma(y') \cap S(x)$. В силу однородности существует слово x' , такое, что $x' \in S(y') \cap \Gamma(x)$. Это означает, что $x' = x$. Итак, $S(x) = S(y)$, и в $\Gamma(y) \cap S(y)$ входят два разных слова y и y' , а, это противоречит требованию простоты. Лемма доказана.

Примечание. Лемма 4 интересна и с точки зрения лингвистической интерпретации. Поскольку слово, входящее в единичную окрестность, по нашей интерпретации есть служебное слово, то мы доказали, что слово, эквивалентное со служебным, само есть служебное слово.

Переходим к доказательству теоремы. Пусть две фразы A и B связаны условиями семантической связи. Мы хотим показать, что любая фраза A' , имеющая ту же структуру, что и A , семантически связана с некоторой фразой B' , имеющей ту же структуру, что и B . При этом нам, очевидно, достаточно доказать, что при замене одного слова, входящего в неединичную окрестность, на произвольное слово из того же семейства соответствующая фраза B' найдется, так как доказательство автоматически распространяется на общий случай при помощи метода полной индукции.

Итак, заменим первое слово из A на произвольное слово из того же семейства и получим таким образом фразу A' . Если это слово входило в единичную окрестность, то, согласно лемме, заменившее его слово входит в единичную окрестность, и в качестве фразы B' , удовлетворяющей условию семантической связи, можно просто взять фразу B . Если же заменяемое слово (обозначим его x_1 , а заменяющее слово v_1) входило в неединичную окрестность, то, во-первых, мы знаем что имелось y_1 во фразе B , такое, что $x_1 \in \Gamma(y_1)$ и, во-вторых, $x_1 \in S(v_1)$. Итак, $x_1 \in \Gamma(y_1) \cap S(v_1)$, а значит, существует (в силу однородности) слово z_1 , такое, что $z_1 \in \Gamma(v_1) \cap S(y_1)$. Поскольку $z_1 \in S(y_1)$, то, заменив в B y_1 на z_1 , мы получим отмеченную фразу B' . Эта фраза семантически связана с A' , поскольку $z_1 \in \Gamma(v_1)$. На основе сделанных выше замечаний можно считать, что теорема доказана.

УКАЗАТЕЛИ

Указатель терминов*

- Аллофон 10
Алломорф 17, 30, 31
Атрибут 43, 51
Архифонема 10
- Базис 41
 базисное упрощение 41
 базисный составляющий 43
 простой б. 41
- Гетеропимия 3
Глубина языка 47
Грамматема 26
Граф 15, 16
- Дерево фразы 46, 47
Динамическая модель 5, 14, 30
Доминирование, 11
- Звук речи 9
- Интонация 21
- Категория 2, 25
 морфологическая к. 25, 26
 парадигматическая к. 25
 равномерно распределенная к. 29
 семантическая к. 23
 элементарная грамматическая к. 25, 29, 36, 51
 элементарная фонологическая к. 10, 11
- Класс
 изолированный к. 12
 контекстуальный к. 23
 морфологический к. 27
 фонологический к. 12
- Конструкция 39, 44, 50
- Конфигурация 39
 к. n-го ранга 39
 предбазисная к. 42
 регулярная к. 42
Кортеж 2
- Логико-грамматические типы предложений 44
- Местоименне 40
Морфема 17, 31
- Нейтрализация
 позиция нейтрализации 10, 26
 н. синтагматическая и парадигматическая 33
- Норма S-структуры 41
- Однородность 4, 13, 26, 28—30
 категориальная о. 26
 признака с.м. фонетический признак
 фонологическая о. 13
 формальная о. 4, 28—30, 48
- Окрестность. 24, 27—31, 48
 правильная о. 31
 подокрестность 29
- Окружение 23
- Омоним
 морфологический о. 32
 синтаксический о. 49
- Оппозиции 12
- Отмеченность 2, 21, 22
- Основа 31
- Падеж 26, 29, 30, 34
Парасинтетизм 19
Пауза 10
План выражения и план содержания 24, 25, 33

* Во всех указателях номера соответствуют параграфам.

- Подкласс морфологический 29
 Подсистема фонологическая 12
 Полнота
 подсистемы 13
 распределения фонем 15
 Предложение 43, 44, 51
 односоставное п. 41
 Признак *см.* фонетический признак
 Прогрессивная конструкция 47
 Производное разбиение 27
 Простота 28
 Пустая клетка 13, 26

 Разбиение 2,12, 22
 Раздел 27
 Рамка 38
 Ранг
 р. фонемы 13
 р. конфигурации 39, 40, 47
 Распределение 15, 23
 дополнительное р. 31
 Регрессивная конструкция 47
 Реятема 43
 Рефлексивность 15
 Род 25, 29
 Ряд 14
 фонемный р. 11
 элементарный грамматический
 р. 32.

 Сегментация 17—20
 Семантическая инвариантность 48,
 49
 Семейство 23, 27—31, 48
 правильное с. 31
 подсемейство 29
 Серия 14
 Синоним синтаксический 49
 Слово 2, 17
 полнозначное с. 43
 служебное с. 43, 48
 фонетическое с., 9
 Словоформа 2
 Совершенная структура *см.* струк-
 тура совершенная
 Совершенство трансформации 48
 Совместимость 10
 Составляющий 39, 43
 непосредственно с. 43
 Спутник 42
 Структура 22
 В-структура 22
 S-структура 23
 отмеченная с. 22
 совершенная с. 22, 39

 Тип 27
 Типология языков 7, 13, 15, 20,
 28, 47—49
 ипоформа 36
 Тождественность смысла 49
 Трансформация 48

 Упрощение данной S-структуры 41
 Устранимость простоты и неод-
 нородности 28

 Флексия 20, 31
 Фонема 9—11
 Фонемный ряд 11
 Фонетический признак 2, 9, 10
 ф. п. однородный с 10
 ф. п. релевантный 10
 ф. п. связанный 10
 ф. п. совместимый с 10
 Фонологическая подсистема 12
 Фраза 2, 21
 отмеченная ф. 21
 ф. с многоточием 23
 ф. с рамкой 38
 Функциональная нагрузка 13

 Часть речи 24, 27, 28
 Число 29, 30
 Член предложения 43

 Эквивалентность 23
 ограниченная э. 45
 практическая э. 45
 э. по рангу 40
 Элемент 2
 Элементарная фонологическая ка-
 тегория *см.* категория
 Элементарный ряд *см.* ряд

 Ядро
 синтаксическое я. языка 48
 я. конфигурации 42
 я. фонологической подсистемы
 13

 Язык
 нормальный 29
 однородный 28
 простой 28
 правильный 27
 примитивный 51
 регулярный 29
 терминальный 46
 с конечным числом состояний
 16, 38
 с полной трансформацией 48

Именной указатель

Аванесов Р. И. 11
Айдукевич К. 23, 51
Андреев Н. Д. 1, 31
Апресян Ю. Д. 52

Балли Ш. 47
Бар-Хиллел И. 8, 44, 51
Бодуэн де Куртене И. А. 13, 52

Василлиу Эм. 15
Винер Н. 52
Виноградов В. В. 24

Гринберг Дж. 3, 20, 22, 31
Гроот де А. 39
Гуссерль Эд. 23

Джоунз Д. 9
Джуз М. 2
Добрушин Р. Л. 25, 32, 35

Ельмслев Л. 15, 33, 34
Есперсен О. 3, 39

Зализняк А. А. 7
Зиддер Л. Р. 1, 10, 16

Иванов Вяч. Вс. 7, 26
Ингве В. 47, 52

Карнап Р. 23
Карцевский С. 21
Колмогоров А. Н. 23
Кузнецов П. С. 9, 20, 26, 30
Кулагина О. С. 22, 23, 27, 28, 29, 39
Курилович Ю. 24

Ляпунов А. А. 22

Марков А. А. 45
Маркус С. 15
Мартине А. 13, 14, 52

Мельчук И. А. 25, 26
Миллер Дж. 47
Молошная Т. Н. 51

Николаева Т. М. 5

Пайк К. 43
Пешковский А. М. 23, 25, 28, 31
Питмен Р. 43

Рассел Б. 2
Реформатский А. А. 19

Сгалл П. 36
Соссюр Ф. 1, 52
Сэпир Эд. 2, 20

Топоров В. Н. 7
Трика Б. 26, 33
Трубецкой Н. С. 2, 10, 12, 15, 18,
24
Тьюринг А. 52

Успенский В. А. 23, 27, 28

Фортунатов Ф. Ф. 30
Фитиалов С. Я. 25, 36

Харрис З. 21, 39, 45
Хинчин А. Я. 1
Хокетт Ч. 3
Холодович А. А. 22
Хомский Н. 3, 21, 38, 46, 47, 52

Цейтин Г. С. 3

Черри К. 12
Чистович Л. А. 5

Шаумян С. К. 2

Щерба Л. В. 3, 10, 23

Якобсон Р. 3, 9, 10, 12, 26
Якубинский Л. П. 9

Указатель языков

английский 26, 47
древнеанглийский 34
армянский 18, 29, 34, 47
болгарский 26, 28, 47, 51
индонезийский 23
литовский 47
немецкий 13, 18, 28
польский 13, 15, 18, 28, 29

русский 9, 10, 11, 12, 13, 15, 19,
26, 28, 29, 41, 44, 47, 48
сербский 15, 28
словацкий 24
суахили 23
финно-угорские 15, 19, 28, 47
финский 19, 28
французский 18, 28
чешский 15, 18, 24, 26, 28, 29, 34,
36, 48
эстонский 15, 28, 47, 48

О Г Л А В Л Е Н И Е

От автора	3
Г л а в а I. Типы моделей языка	
§ 1. Дедуктивные методы в лингвистике	6
§ 2. Модель и интерпретация. Общий вид модели языка	9
§ 3. Аналитические и синтетические модели	12
§ 4. Распознающие модели	15
§ 5. Парадигматические и синтагматические модели	16
§ 6. Модели языка и статистика речи.	16
§ 7. Теория моделей и структурная типология языков	17
§ 8. Теория моделей как лингвистическая теория	18
Г л а в а II. Методы моделирования в фонологии	
§ 9. Основные понятия при построении фонологических моделей и их лингвистическая интерпретация	20
§ 10. Две классификации фонетических признаков. Фонема и архифонема	23
§ 11. Фонема в московской фонологической школе и ее формальный аналог	25
§ 12. Парадигматическая модель	26
§ 13. Фонологическая однородность. Понятие пустых клеток	31
§ 14. Динамическая модель Мартине	35
§ 15. Синтагматические модели	37
§ 16. Модель порождения	50
Г л а в а III. Основные понятия при построении грамматических моделей и их лингвистическая интерпретация	
§ 17. Морфема или слово?	53
§ 18. Фонологические критерии сегментации	54
§ 19. Конструктивный критерий сегментации	55
§ 20. Основные речевые единицы	56
§ 21. Понятие отмеченной фразы	60
§ 22. Понятие В-структуры фразы	62
§ 23. Понятие эквивалентности	63
§ 24. Понятие окрестности	69
§ 25. Грамматические категории	72
§ 26. О так называемом изоморфизме между фонологией и грамматикой	75
Г л а в а IV. Парадигматические модели в грамматике	
§ 27. Моделирование понятия «часть речи»	81
§ 28. Простота и однородность языка	85
§ 29. Подокрестность и подсемейство. Связь между подсемейством и элементарной категорией	91

§ 30. Вопрос о разграничении индивидуальных парадигм . . .	97
§ 31. Замечания относительно морфемного анализа	101
§ 32. Проблема грамматической омонимии	105
§ 33. О связи между синкретизмом и нейтрализацией	107
§ 34. Соответствие между грамматическими категориями и элементарными рядами	108
§ 35. Расширение модели элементарного ряда	112
§ 36. Замечания о построении распознающей модели	113

Глава V. Синтагматические модели в грамматике

§ 37. Модель, перенесенная из фонологии	116
§ 38. Язык с конечным числом состояний	117
§ 39. Конфигурационная модель Кулагиной	122
§ 40. Эквивалентность по рангу	125
§ 41. Норма S-структуры.	126
§ 42. Регулярные и нерегулярные конфигурации	127
§ 43. Составляющие и отношения между ними. Предложение	128
§ 44. Конструкции	134
§ 45. Должно ли множество отмеченных фраз быть конечным или бесконечным?	135
§ 46. Порождение методом развертывания	137
§ 47. Гипотеза Ингве	140
§ 48. Трансформационный анализ.	145
§ 49. О составлении списка трансформаций применительно к конкретным языкам	149
§ 50. Трансформационный синтез	152
§ 51. К построению распознающей модели	154
§ 52. Модель порождения и парадигматическая модель.	161
Приложение	168
Указатель терминов	187
Именной указатель	189
Указатель языков	189

Исаак Иосифович Ревзин

Модели языка

*Утверждено к печати
Институтом славяноведения
Академии наук СССР*

Редактор Издательства *И. Ю. Шехтер*
Технические редакторы *Т. А. Прусакова и А. П. Гусева*
Художник *М. И. Эльцуфен*
Корректор *В. Т. Макаров*

РИСО-АН СССР № 86-85В. Сдано в набор 25/1 1962 г.
Подписано к печати 11/VI 1962 г. Формат 60×92¹/₁₆.
Печ. л. 12. Уч.-изд. л. 11,2. Тираж 4000.
Т-05821. Изд. № 534. Тип. зак. 309

Цена 77 к.

Издательство Академии наук СССР.
Москва, Б-62, Подсосенский пер., 21

2-я типография Издательства АН СССР.
Москва, Г-99, Шубинский пер., 10

