С.-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ПРИКЛАДНОЕ ЯЗЫКОЗНАНИЕ

УЧЕБНИК

Отв. редактор *А. С. Герд*

САНКТ-ПЕТЕРБУРГ
ИЗДАТЕЛЬСТВО С.-ПЕТЕРБУРГСКОГО УНИВЕРСИТЕТА
1996

ББК 81.1-923
П75

Авторы: С. А. Аверина, И. В. Азарова, Е. Л. Алексеева, Л. Н. Беляева, В. В. Богданов, Л. В. Бондарко, В. Д. Буторов, Л. А. Вербицкая, В. Г. Ветвицкий, А. С. Герд, О. Н. Гринбаум, И. В. Жарков, В. П. Захаров, Л. Р. Зиндер, Т. И. Зубкова, В. Ф. Иванова, Р. Ю. Кобрин, М. А. Красноперова, В. М. Лейчик, Г. Я. Мартыненко, М. А. Марусенко, М. И. Откупщикова, И. П. Панков, З. В. Партыко, Е. Д. Савенкова, И. И. Убин, С. В. Чебанов, А. С. Штерн

Отв. редактор А. С. Герд

Рецензенты: д-р филол. наук В. Б. Касевич (С.-Петербург. ун-т), д-р филол. наук Л. П. Ступин (С.-Петербург. ун-т)

Печатается по постановлению
Редакционно-издательского совета
С.-Петербургского университета

Прикладное языкознание: Учебник / Л. В. Бондарко,
П75 Л. А. Вербицкая, Г. Я. Мартыненко и др.; Отв. редактор А. С. Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. — 528 с.
ISBN 5-288-01369-1

Это первый в нашей стране фундаментальный учебник для филологических факультетов университетов и педагогических вузов. Основное внимание уделяется собственно гуманитарным аспектам применения ЭВМ в филологии, что позволяет использовать данное издание в преподавании филологических дисциплин и при чтении различных курсов и спецкурсов в вузах, школах и лицеях. Освещаются как традиционные, так и новейшие проблемы прикладного языкознания: искусственный интеллект, машинный перевод, разработка информационных систем и др.

Для студентов филологических факультетов университетов и педагогических институтов и всех интересующихся математической и прикладной лингвистикой, вопросами применения ЭВМ в гуманитарных областях.

П $\frac{4602010000-100}{076(02)-96}$ 95—94(177—96)

ББК 81.1-923

ISBN 5-288-01369-1

© Л. В. Бондарко, Л. А. Вербицкая, Г. Я. Мартыненко и др., 1996
© Издательство С.-Петербургского университета, 1996

ОТ РЕДАКТОРА

Настоящий учебник прошел долгий и нелегкий путь. Он зародился в конце 70-х — начале 80-х годов в виде министерской программы спецкурса по новым аспектам прикладной лингвистики, написанной В. В. Богдановым; редакторами этой программы были Р. Г. Пиотровский и Б. Н. Головин.

К 1983 г. при кафедре математической лингвистики Ленинградского университета оформился план учебника широкого профиля. Шли годы, рос авторский коллектив, уточнялись аспекты. Время отсеяло некоторые чисто прагматические, сугубо временные задачи, ярче высветило основные проблемы и направления прикладного языкознания.

Главная особенность созданного нами учебника состоит в том, что он, в отличие от некоторых уже имеющихся, основное внимание уделяет не процедурным и технологическим аспектам работы с техническими средствами (ЭВМ, подготовка, работа и общение с ЭВМ, обработка данных на ЭВМ), а прикладным проблемам общефилологического и даже общегуманитарного характера и современным методам их решения.

Цель данного учебника — системное изложение проблем и методов современной прикладной лингвистики. В его задачу не входит исчерпывающий охват всех проблем, с которыми в настоящее время приходится иметь дело прикладной лингвистике. Авторы считают необходимым и целесообразным ограничиться изложением тех проблем, которые, по их мнению, являются актуальными в свете современных требований науки, экономики и техники, предъявляемых к лингвистике, а также показать применение методов прикладной лингвистики в различных гуманитарных областях.

В условиях научно-технической революции, сопровождающейся бурным ростом информационных массивов и интенсифи-

кацией информационных потоков, выдвижение на передний план проблематики, связанной с автоматической переработкой информации и информационным поиском, представляется вполне разумным и оправданным шагом.

В одно учебное пособие не могли быть помещены главы о моделировании на всех уровнях языка, отражены различные аспекты теории связного текста, все конкретные прикладные аспекты; не освещена также специфика применения новых методов и автоматизации в диалектологии, ономастике. Во многом она выводима из общих глав учебника. Не отражены в полном виде и проблемы моделирования знаний, методики преподавания неродного языка, языковой политики. Одной из важных задач прикладной лингвистики является создание машинных фондов отдельных языков.

При каждой главе приводится лишь самая основная литература, пользуясь которой читатель сможет найти уже и работы более специальные.

Все авторы данного учебника в течение многих лет принимают активное участие в выполнении самых разных конкретных прикладных тем, и в этом отношении настоящая книга отражает опыт, накопленный ими в ходе разработки таких проблем. В то же время многоаспектность проблематики, большой авторский коллектив обусловили и некоторую структурную неоднородность отдельных глав.

Естественно, что, как и всякий подлинно университетский учебник, он вырос на основе тех лекционных курсов, которые читали его авторы.

А. С. Герд

ПРЕДМЕТ И ОСНОВНЫЕ НАПРАВЛЕНИЯ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

Прикладные сферы языкознания издавна отличались широким разнообразием. Наиболее древние из них — письмо (графика), методика обучения родному и неродному языкам, лексикография. В дальнейшем появились перевод, дешифровка, орфография, транслитерация, разработка терминологии. Одно из традиционных направлений прикладной лингвистики — участие в языковой политике государства. Языковая политика включает в себя: 1) разработку алфавита, принципов орфографии, ликвидацию неграмотности; 2) выбор и установление государственного языка; 3) определение положения других языков по отношению к языку государственному; 4) разработку, унификацию и стандартизацию национальной терминологии; 5) вопросы именования и переименования (городов, улиц, площадей).

Одновременно с развитием и совершенствованием классических прикладных областей языкознания во второй половине XX в. наметился и определился ряд новых направлений прикладной лингвистики. Эти направления представляют собой логическое продолжение исторического развития прикладной лингвистики на новом этапе, обусловленном современной научно-технической революцией, характеризующейся укреплением взаимосвязи общественных, естественных и технических наук.

В целом прикладные аспекты лингвистического обеспечения разнообразных сфер человеческой деятельности сводятся прежде всего к одной общей проблеме — проблеме обработки информации, функционирующей в обществе. Это и текстовая информация в ее письменном виде, и устная речь как наиболее привычный способ коммуникации. Особая роль языкознания в решении практических проблем и потребностей общества определяется самой сущностью естественного человеческого языка, являющегося уникальным средством хранения и передачи информации.

Актуальные практические проблемы и потребности общества — это внедрение в производство станков и приборов, управляемых голосом оператора, систем автоматизированного поиска и обработки информации на разных языках, совершенствование организации связи (телефонной, радиосвязи и др.), использование фонетических данных при лечении нарушений речи и слуха.

Названные проблемы включают в себя ряд новых направлений прикладной лингвистики, связанных с автоматической обработкой текста (речи), с широким использованием ЭВМ, таких, как лингвистическое обеспечение информационных систем разных типов; машинный перевод; компьютеризация обучения; разработка систем, понимающих естественный язык (лингвистические задачи в системах искусственного интеллекта); разработка систем использования информации, содержащейся в звуковом речевом сигнале.

В связи с разработкой лингвистического обеспечения информационных систем разных типов в последние годы особую актуальность приобрела проблема унификации и стандартизации терминологии. Это связано с резко возросшим количеством научных и технических терминов, обслуживающих различные сферы знания. Естественно, что при этом не остается в стороне и лексикография, для которой вопросы, связанные со стандартизацией терминологии, являются в высшей степени животрепещущими. Постепенно из традиционной лексикографии все определеннее выделяется специальная сфера лексикографической деятельности, связанная с широким применением ЭВМ, — автоматическая лексикография.

При решении ряда практических задач требуется эффективное взаимодействие человека с ЭВМ в процессе диалога на естественном языке. Таковы задачи разработки вопросно-ответных систем, управления роботом, обеспечения диалога в процессе принятия решений в системе управления. Основная проблема здесь состоит в том, чтобы обеспечить понимание письменного текста или звучащей речи (если используется устный ввод) автоматическими устройствами.

Решение каждой новой прикладной задачи не тривиально и не шаблонно. Число и виды прикладных лингвистических задач непредсказуемы, так как нельзя предвидеть всех коммуникативных ситуаций типа «человек — человек» и «человек — ЭВМ», в разрешении которых может понадобиться участие лингвистов; сами задачи меняются: одни уходят, приходят новые.

Основной корпус фундаментальных исследований, проводимых в недрах прикладной лингвистики, в конечном счете, составляют описание и моделирование фонетической, грамматической, семантической и статистической структуры научных,

технических и организационно-распорядительных устно-речевых и письменных документов, создание словарей и семантических представлений.

В частности, построение формальных моделей текстовых данных является, по существу, уже пограничной областью между собственно теоретической лингвистикой и прикладными исследованиями. Сам тип такой модели указывает обычно и пути ее алгоритмизации.

Прикладная лингвистика возникла в результате острой потребности знать, что такое язык в действии, в коммуникации, поскольку теоретическая лингвистика изучала преимущественно язык в его состоянии, в системе. При этом она долго чуждалась и сторонилась своего родства с научно-технической и деловой прозой. Не было описаний, моделей, ориентированных на собственно прикладные задачи. Наконец, в 70-х годах XX в. ученые пришли к выводу, что решение многих прикладных проблем не может быть чисто лингвистическим, а лежит на совсем иных путях, на путях моделирования поведения и мышления человека, семантики, синтеза формальных и семантических средств языка. Так появилась одна из важнейших межотраслевых фундаментальных проблем прикладной направленности — проблема моделирования знаний.

Моделирование знаний в последнее время оказалось сферой приложений самых разных наук: логики, лингвистики, математики, психологии, кибернетики. Причина такого всеобщего интереса к этой проблеме одна — создание автоматических систем искусственного интеллекта. В то же время знания нам даны в языке и через язык. Знания существуют как в текстах спонтанной речи (монолог, диалог, реплики в разных ситуациях), так и в письменных текстах (памятники древней письменности, художественная литература, научная и техническая литература). Через язык мы передаем свои знания от поколения к поколению. Язык, таким образом, выступает и как форма хранения знаний, и как средство и способ их передачи. В науке и технике, по-видимому, вообще нет знаний вне текстов, а моделирование семантики научно-технических текстов — это моделирование системы знаний этой отрасли. Так постепенно проблема моделирования знаний сомкнулась с проблемой моделирования смысла текста. И здесь центральным является вопрос о создании семантических представлений текста.

В то же время вопрос о том, как решать ту или иную конкретную прикладную задачу в сфере технической и деловой коммуникации, определяется именно тем, насколько детально описаны тексты этих документов, их структура, грамматика, лексика, семантика, есть ли терминологический словарь, выявлена ли статистическая структура текста, построены ли семантические представления для текстов этого типа.

В свою очередь, усиление темпов эволюции прикладного языкознания уже не раз порождало ускоренное развитие и самой лингвистической теории. Так, в 20—30-х годах XX в. практические потребности унификации и стандартизации научно-технической терминологии привели к появлению новой лингвистической дисциплины — терминоведения. Грандиозная по своим масштабам практическая работа по созданию алфавитов и письменности для языков народов СССР стимулировала в 30—40-х годах совершенствование методов синхронического описания языков.

Все новые аспекты прикладного языкознания, появившиеся за последние 20—30 лет, связаны с одной общей проблемой — автоматической обработкой языка в его письменной и устной формах. Возникновение этой проблемы привело к появлению новых методов анализа языка, новых приемов его описания, способствовало развитию новых взглядов на природу языка и на структуру языкознания в целом.

Решение сложной проблемы автоматической переработки языкового материала в 50—70-х годах XX в. во многом продвинуло и дальнейшее развитие лингвистической теории.

Новые аспекты прикладного языкознания ускорили начавшееся проникновение в теоретическое языкознание методов математики, в особенности методов теоретико-множественных, формально-логических, статистико-вероятностных.

На пересечении методов классического языкознания, классической логики, психологии, семиотики и математики выросла современная структурная лингвистика как целостная наука о методах моделирования в языкознании. В лингвистике появился и оформился новый самостоятельный теоретический раздел — теория моделей языка.

На границе между лингвистикой и математикой выросла новая дисциплина — математическая лингвистика. Возникновение и оформление ее в 50—60-х годах XX в. было уже давно подготовлено всем внутренним развитием языкознания. Внимание к структуре языка, подход к языку как к системе, осознание языковой системы как сложной структуры, состоящей из микросистем, выявление разных типов отношений между лингвистическими объектами и их атрибутами — все это объективно создало те условия, в которых математическая лингвистика не могла не возникнуть.

Именно эти две дисциплины — структурная лингвистика и математическая лингвистика — оказали наибольшее влияние на развитие современного теоретического языкознания. Сильнее всего это влияние сказалось на синтаксисе и семантике.

В структурном синтаксисе были разработаны две формальные синтаксические модели — модель непосредственно составляющих и модель зависимостей. Обе модели широко применяются при автоматическом синтаксическом анализе, в машин-

ном переводе и в других задачах автоматической обработки текста; используются также «гибридные» синтаксические представления, учитывающие достоинства той или иной модели.

Семантика пронизывает все уровни языка и тем самым не представляет собой отдельного уровня. В семантике постоянно сказывается влияние, с одной стороны, методов собственно лингвистических, а с другой — методов логической семантики, теории классификаций.

При моделировании лексической семантики в структурной лингвистике возникли теории семантического поля, компонентного анализа, тезаурусного описания лексики. В синтаксической семантике было разработано большое число концепций семантической структуры предложения.

Многие теоретические понятия и методы исследования и анализа, появившиеся среди гуманитарных наук впервые в структурной, математической и прикладной лингвистике, сегодня все более активно проникают в литературоведение, библиотековедение, историю, этнографию, психологию, документоведение. В различных гуманитарных науках много общего и в направленности автоматизации исследований.

Автоматизация научных исследований в гуманитарных науках должна идти прежде всего по линии создания больших банков хранения документальных и фактографических данных. Такой тип автоматизации продиктован прежде всего характером основных информационных потребностей специалистов по гуманитарным наукам.

Какого рода справки чаще всего нужны научному работнику-гуманитарию? Во-первых, это запросы библиотечно-библиографического типа: найти и выдать всю отечественную и зарубежную литературу (книги, статьи, тезисы конференций, хроникальные заметки) по отдельному вопросу. Например, археологу нужна литература о височных украшениях определенного вида, этнографу — о типах хозяйственных построек в Северной Европе, лингвисту — о сочетаемости морфем.

Во-вторых, в археологии, истории, этнографии антропологии, источниковедении, языкознании в ходе исследования чаще всего бывает нужно произвести подбор как опубликованных, так и неопубликованных источников, содержащих материалы того или иного типа. Работа эта, как правило, крайне трудоемка. Археологу нужно, например, собрать все конкретные описания предметов текстильной керамики, этнографу — отыскать все источники, содержащие описание орудий рыболовства в Восточной Европе, филологу — все жития из разных описаний рукописей и все книги, содержащие публикацию таких памятников; все текстовые источники, отражающие, например, конфессионально-повествовательный тип украинских текстов XVI в., историку — все работы о торговле русских монастырей в XVI—XVII вв. Подбор источников в фундаментальных гуманитарных

исследованиях не знает ни хронологических, ни географических, ни языковых границ. Более того, нередко крайне ценная, собственно фактографическая информация содержится именно в старых источниках. Ср., например, путешествия русских академиков XVIII в. Гмелина, Палласа, Озерецковского, Георги, Лепехина, бесценные статистические отчеты по губерниям России, издававшиеся в XIX — начале XX в., или писцовые, приходо-расходные, таможенные книги XVI — XVII вв., списки населенных мест Российской империи.

В-третьих, перед исследователем-гуманитарием стоит задача отбора из источников «своих» фактов из моря другого материала. Так, например, археологу из огромного количества изделий из керамики, собранных в XIX—XX вв., нужно отобрать только объекты «своего» рассмотрения, например, керамику именно с отпечатками ткани, этнографу — сведения о деревянной посуде, филологу — слова с суффиксом *-ость* или глаголы прошедшего времени. В археологии, этнографии, антропологии, частично в лингвистике множество данных, конкретных фактов, полученных в поле, в экспедициях, хранится обычно в неопубликованном виде — в архивах, фондах музеев, институтов, вузов, в ручных описаниях и картотеках. Неопубликованные материалы такого рода, разбросанные к тому же по разным краям и городам, — бесценный источник любого научного исследования в области историко-филологических наук в целом. Определить тип, характер и генезис того или иного явления в гуманитарных науках можно только на широком сравнительно-историческом и типологическом фоне.

Особое значение приобретают очерченные выше проблемы в междисциплинарных, комплексных исследованиях. Так, например, современная историческая география оперирует одновременно фактами из физической и экономической географии, археологии, истории, этнографии, демографии. Специалисту по теории этногенеза и этнической истории нужны сведения из археологии, антропологии, языкознания, источниковедения, палеозоологии, палеоботаники, музыковедения, фольклористики. Решение вопросов этнической истории и исторической географии Восточной Европы требует обращения не только к фактам археологии, к ареалу и генезису отдельных культур, но и к сочинениям Иордана, Геродота, Прокопия Кесарийского, Тацита, к сопоставлению древних текстов, анализу старых карт, к данным языкознания (топонимики). А как известно, именно на стыке разных наук и появляются наиболее интересные решения и гипотезы.

Наконец, перед исследователем в области гуманитарных наук стоит задача систематизации собранного материала. При этом для большинства гуманитарных исследований, как правило, бывает нужна классификация данных по разным параметрам. Например, в археологии: типы погребений могут быть

охарактеризованы по их структуре, по хронологии, по ареалу; аналогично в этнографии: виды построек — по материалу, планировке, назначению, по ареалу; в языкознании: словоформы — по внутренней структуре, по их отношению к системе языка, по их реализации в речи, в текстах разных жанров, по носителям языка, по хронологии, по ареалу, по школам письменности. Все это предполагает неоднократную переклассификацию одного и того же материала по разным параметрам.

Все сказанное предопределяет и специфические первичные формы представления фактов в гуманитарных науках. В любой из них большое значение имеет создание капитальных сводов, картотек, каталогов, реестров и индексов фактов, словарей, содержащих описание данных и их детальную характеристику. Создание таких сводов во многом представляет собой самостоятельную подлинно исследовательскую задачу. Таковы описания и каталоги архивных музейных собраний, картотеки слов и форм в лингвистике, описания рукописей библиотек, перечни названий племен и народов, фамилий, географических названий, прикнижные словари и предметные указатели (см., например, индексы в многотомном издании «Русская историческая библиотека» и др.).

Создание таких сводов, картотек, словарей отнюдь не механическое дело, как это иногда представляется со стороны, ибо каждый объект снабжается описанием всех его содержательных параметров, релевантных для специалистов. При этом сами описания строятся на разных принципах — по алфавиту, тематически, хронологически, по ареалу. Проблема типов вторичных описаний — особая область исследований. Для всех гуманитарных наук, в отличие от отраслей научно-технических, характерно нестарение самих фактов, данных. Вещи, найденные в раскопках XIX в., описания быта русских крестьян конца XVIII в., записи речи, сделанные в середине XIX в., имеют такую же, а порой и бо́льшую, ценность, как и материалы, обнаруженные, полученные в наши дни. Более того, для истории, этнографии, языкознания трудно переоценить, например, значение материалов XV—XVII вв., так как в этих данных зафиксировано состояние, которое сегодня уже исчезло.

Разработка многих гуманитарных проблем, в особенности языкознания, этнографии, антропологии, истории, связана с привлечением карт как источников. Вот почему проблема ввода в ЭВМ и вывода данных карт представляется чрезвычайно актуальной не только для географии, геологии или биологии, но и для гуманитарных исследований. Более того, именно обработка картографических данных и построение новых карт представляет собой один из наиболее трудоемких этапов такой работы.

Наконец, этап собственно исследовательского анализа материала в гуманитарных науках также, как правило, невозмо-

жен без постоянного обращения к литературе прошлого, к мнениям отечественных и зарубежных ученых, по крайней мере с середины XVIII в.

Автоматизация научно-исследовательских работ в гуманитарных науках должна оптимизировать:

а) поиск литературы предмета на разных языках;

б) подбор источников, фондов материалов;

в) оперирование данными источников;

г) многоаспектную, глубоко эшелонированную классификацию материала;

д) создание сводных описаний, реестров, каталогов по заранее заданным параметрам;

е) применение методов статистики, картографии, теории классификации и системного анализа;

ж) графическое представление данных в виде схем, рисунков, карт.

Таким образом, в процессе автоматизации научных работ система должна выдавать специалисту максимально полные на данном историческом уровне энциклопедические знания и библиографические сведения об объекте.

Ядром современного прикладного языкознания является структурная и математическая лингвистика во всем многообразии их методов и достижений.

Основной предмет структурной и математической лингвистики — разработка и совершенствование структурных и формальных методов анализа и описания языка. В наши дни многие, хотя и не все, приемы и достижения структурной лингвистики стали уже достоянием общего языкознания.

В целом, во всей своей совокупности целей, направлений, методов и подходов, прикладная лингвистика — это прежде всего комплексная дисциплина, постоянно соприкасающаяся с философией, психологией, физиологией, математикой, логикой, социологией, информатикой.

Прилагаемый ниже список основных терминов (понятий) математической и прикладной лингвистики дает определенное представление о предмете, направлениях, методах этой науки сегодня.

Алгоритм — последовательность действий, выполнение которых необходимо для достижения конкретной цели.

База данных — совокупность данных, представленных в формализованном виде, удобном для автоматической передачи, интерпретации или обработки.

База знаний — база данных, элементами которых являются формализованные представления знаний об элементах внеязыковой действительности.

Байт — в вычислительной технике — обрабатываемый как единое целое элемент данных, представляющий собой последовательность двоичных разрядов (обычно — 8). Используется и как единица измерения памяти. 1 килобайт=1024 байта. 1 мегабайт=1024 килобайта. При обработке и хранении текстов обычно одному байту соответствует один символ текста.

Грамматика зависимостей — формальное представление строя предложения в виде иерархии компонентов, между которыми установлено отношение зависимости.

Грамматика непосредственно составляющих — формальное представление строя предложения в виде иерархии вложенных друг в друга линейно непересекающихся элементов, максимально независимых друг от друга.

Грамматика представлений — система правил, основанная на учете возможных окружений каждого слова.

Граф — математический объект, состоящий из множества вершин (точек) и множества ребер (связей), соединяющих пары вершин.

Дерево предложения, дерево зависимостей — представление структуры предложения в виде ориентированного графа, узлы которого соответствуют компонентам структуры зависимостей.

Интерфейс — средства и способы обмена информации. Обычно имеют в виду человеко-машинный интерфейс, т. е. средства и способы организации диалога человека и ЭВМ, хотя можно говорить и об интерфейсе между отдельными блоками вычислительной машины, программами, частями программ и т. д. Одной из лингвистических задач искусственного интеллекта является создание естественно-языкового интерфейса, т. е. организация диалога человека с ЭВМ на естественном языке.

Квазиреферат — реферат, составленный компьютером.

Конфигурационный анализ — анализ входного текста путем сравнения с набором предварительно выделенных синтаксических конструкций (конфигураций, формул) переводимого языка. При совпадении установленной конфигурации входного текста с одной из конфигураций набора текстовая конфигурация считается распознанной, «свертывается» и в дальнейшем анализе фигурирует в свернутом виде. При синтезе полученная из анализа конфигурация сравнивается с соответствующей ей конфигурацией переводящего языка. Соответствие конфигураций устанавливается предварительно.

Лингвистический процессор — набор процедур для обработки текстов на естественном языке, которые являются входной информацией автоматизированных систем.

Накопитель — устройство в ЭВМ, предназначенное для считывания и записи информации на внешних носителях. Например, накопитель на магнитных дисках (дисковод), накопитель на магнитных лентах (лентопротяжное устройство) и т. п.

Оконный интерфейс — один из современных способов организации интерфейса (см.). При этом способе на экране дисплея выделяется прямоугольная область (окно), в которую человек вводит и/или ЭВМ выводит информацию. При необходимости организации диалога на различные темы может быть организовано несколько окон (многооконный интерфейс).

Предсказуемостный анализ — процедура, основанная на просмотре текста слева направо, при этом для каждого очередного слова проверяется, является ли оно по тем или иным параметрам предсказанным предыдущими.

Псевдограф — ориентированный граф, имеющий и петли, и дуги.

Семантическое представление данных — формальная запись смысловой интерпретации информации, которая использует какую-либо модель семантического описания.

Словарь-конкорданс — специальный словарь текстовых словоформ с указанием всех контекстов для каждой словоформы.

Тезаурус — идеографический словарь, в котором показаны семантические отношения между его единицами. Структурной основой тезауруса является иерархическая система понятий конкретной предметной области.

Файл — совокупность связанных записей в памяти ЭВМ, рассматриваемая как единое целое.

Фрейм — способ описания структуры текста или ситуации, состоящий в перечислении возможных компонентов и типов связей между ними.

Язык представления данных — формальный способ структурирования за-

писи информации, в основе которого лежит какая-либо формализованная (математическая) модель описания.*

Литература

Бектаев К. Б., Пиотровский Р. Г. Математические методы в языкознании. Ч. I. Алма-Ата, 1973; ч. II. Алма-Ата, 1974.

Беляева Л. Н. Применение ЭВМ в лингвистических исследованиях и лингводидактике. Л., 1986.

Винер Н. Кибернетика. М., 1983.

Звегинцев В. А. 1) История языкознания XIX—XX веков в очерках и извлечениях. Ч. II. М., 1965; 2) Теоретическая и прикладная лингвистика. М., 1968.

Знание языка и языкознание. М., 1991.

Зубов А. В. Основы лингвистической информатики: Учеб. пособие. Ч. I. Минск, 1991; ч. II. 1992; ч. III. 1993.

Зубов А. В., Лихтарович А. А. ЭВМ анализирует текст. Минск, 1989.

Использование ЭВМ в лингвистических исследованиях / Под ред. В. И. Перебейнос. Киев, 1990.

Компьютеризация общества и человеческий фактор / Под ред. А. И. Ракитова. М., 1988.

Кондратов А. М. Электронный разум. М., 1987.

Котов Р. Г., Якушин Б. В. Языки информационных систем. М., 1979.

Лингвистический энциклопедический словарь. М., 1990.

Марчук Ю. Н. Математические методы в языкознании. М., 1990.

Налимов В. В. Вероятностная модель языка. М., 1974.

Новое в зарубежной лингвистике. Вып. XXIV: Компьютерная лингвистика. М., 1989.

Пиотровский Р. Г. 1) Текст, машина, человек. Л., 1978; 2) Инженерная лингвистика и теория языка. Л., 1979.

Прикладные аспекты лингвистики. М., 1989.

Рогожникова Р. П., Чернышева Л. В., Кузнецова Е. Основные направления автоматизации лингвистических исследований. Л., 1988.

Рождественский Ю. В. Техника, культура, язык. М., 1993.

Структурная и прикладная лингвистика. Вып. I / Под ред. А. С. Герда. Л., 1978.

Шингарева Е. А. Семиотические основы лингвистической информатики. Л., 1987.

* В создании списка терминов принимали участие все авторы.

ВВЕДЕНИЕ В ТЕОРИЮ ПИСЬМА

Огромное значение письменной формы речи для социального прогресса, для развития техники и науки очевидно. Без письменности было бы немыслимо то огромное накопление знаний в различных областях проникновения человека в законы природы и общества, которое характеризует современные высокоразвитые народы. Поэтому не будет преувеличением сказать, что создание письма для бесписьменных языков и совершенствование письма старописьменных языков является важнейшей прикладной задачей языковедения.

Требования практики нередко бывают стимулом для возникновения и развития какой-нибудь науки или научного направления. Так случилось и с языковедением, которое возникло как учение о правильном письме. В античности оно и называлось соответственно грамматикой, от греческого слова γράμματα 'знаки письма, письмена'.

Несмотря на громадный многовековой опыт по созданию письма для многих языков сколько-нибудь развитой общей теории письма не существует до сих пор. Все внимание языковедов было направлено в этой области на историю возникновения и эволюцию письма. Попытки создания общей теории появились только в последние десятилетия.

Каковы же общелингвистические предпосылки такой теории? По словам одного из создателей нового языковедения Вильгельма фон Гумбольдта, «языковые знаки — это обязательно звуки» (Гумбольдт, 1984, с. 302). Специально о роли письма он не писал, но многократно упоминал письмо как средство передачи акустико-звукового облика слова.

Фердинанд де Соссюр образно охарактеризовал значение письма следующими словами: «Язык и письмо суть две различные системы знаков; единственный смысл второй из них — служить для изображения первой; предметом лингвистики яв-

ляется 〈...〉 исключительно звучащее слово. Но графическое слово присваивает себе 〈...〉 столько же или даже больше значения, нежели самому этому знаку. Это все равно, как если бы утверждали, будто для ознакомления с человеком полезнее увидеть его фотографию, нежели его лицо» (Соссюр, 1977, с. 62—63).

В русском языковедении вторичность письма давно считается аксиомой; достаточно сослаться на название книги И. А. Бодуэна де Куртенэ «Об отношении русского письма к русскому языку», вышедшей в свет еще в 1912 г. Из этого названия явствует, что, говоря о языке, мы имеем в виду его звуковую форму.

Очевидность примата языка по отношению к письму вытекает из того, что все люди в детстве сначала неосознанно (спонтанно) овладевают языком, а затем учатся грамоте или же на всю жизнь остаются не знающими письма. Несмотря на все сказанное некоторые лингвисты, искажая истинное положение вещей, утверждают, будто существуют две формы языка (звуковая и оптическая). Письмо же не некая форма существования языка, а лишь оптический код языка, звукового по природе.

В зависимости от того, какая языковая единица является объектом знака письма, различают типы письма. Если таким объектом служит значимая единица языка как некое целое, ее план содержания, то такое письмо называют идеографическим или иероглифическим, а знак такого письма — иероглифом или идеограммой. Письмо, в котором знак имеет своим объектом звуковые единицы (экспоненты значимых единиц), называют звуковым или фонетическим, а знаки такого письма — слоговыми знаками и буквами.

Иероглиф служит знаком для понятия, отраженного в том или ином слове, звуковая же сторона слова в нем совершенно не отражена; так, например, иероглиф, имеющий значение 'дерево', может читаться по-разному в зависимости от того, как слово звучит в данном языке или диалекте.

Своеобразные иероглифы мы имеем в цифрах, в других математических знаках, а также в разных символах, принятых в других науках. Знаки 2, 3, +, —, % и т. п. в тех языках, где они приняты, имеют одно и то же значение, но читаются они по-разному — в зависимости от того, какое слово стоит за ними в данном языке. Цифра 2, например, соответствует русскому *два*, французскому deux, английскому two, немецкому zwei и т. д. Можно сказать, что подобные знаки отличаются от древнеегипетских, китайских и других иероглифов только тем, что они не восходят к рисункам.

Буквы служат знаками для звуковых единиц. Фонетическое письмо отражает только план выражения; поэтому и совершенно незнакомое слово можно прочесть, если известны бук-

вы и графические правила данного языка. Это бывает достаточно часто. Даже при чтении текстов на родном языке нередко встречаются новые, не известные читающему слова; он легко «озвучивает» их, хотя и не понимает.

В большинстве языков мира господствует фонетический тип письма. Знаки иероглифического типа (цифры и т. п.) выполняют вспомогательную функцию, и только в очень специальных текстах они занимают значительное место. Как правило, такие знаки имеют интернациональный характер.

В первые два десятилетия после Октябрьской революции в нашей стране была создана письменность для большого числа бесписьменных языков народов Севера, Сибири, Кавказа и др. Для организации этого дела в начале 20-х годов при Президиуме ВЦИК был создан Центральный комитет нового алфавита, при котором вначале состояла Научная ассоциация, а затем был организован Центральный институт языка и письменности, слившийся в конце 30-х годов с Институтом языка и мышления АН СССР. Также при Президиуме ВЦИК состоял и Комитет нового алфавита народов Севера, тесно связанный с Институтом народов Севера, в котором обучались представители различных народностей Севера и велись интенсивные научные исследования. Активное участие во всем этом принимал Ленинградский университет в лице Л. В. Щербы, его сотрудников и учеников.

К моменту начала деятельности всех этих учреждений в русской науке уже существовала развитая теория фонемы, о которой тогда только начинали говорить на Западе. Знакомство с этой теорией сыграло большую роль при создании алфавитов и правил письма. Благодаря этому письмо создавалось как ф о н е м н о е, а не просто звуковое. Это значит, что за единицу, подлежащую передаче на письме, следует принять фонему, т. е. единицу, в конечном счете связанную со смыслом, а не звук речи как артикуляторно-акустическое явление, представляющий фонему в соответствующей позиции, т. е. аллофон (вариант) фонемы.

Переходя к краткому изложению общей теории языка, следует сказать, что при создании алфавитов для бесписьменных языков многие лингвисты, к сожалению, были мало знакомы с упомянутой выше книгой И. А. Бодуэна де Куртенэ, а также и с другими его трудами, в которых он защищает положение о том, что учение о письме — это часть фонетики. Прежде чем создавать алфавит для того или иного языка, надо выявить его систему фонем; а для создания орфографии, т. е. правил написания слов соответствующего языка, надо определить, какие фонемы образуют звуковой облик данного слова.

При решении последней задачи необходимо помнить, что в спонтанной речи происходит разнообразная редукция отдельных частей слова, которая ведет если не к полной утрате от-

919229

дельных фонем, то к так называемой фонетической неопределенности, т. е. к возможности различной фонемной трактовки редуцированного сегмента. Ярким примером этого могут служить неударенные гласные русского языка; так, последний гласный в слове *сегодня,* которое в фонетической (не фонематической) транскрипции может быть передано как |s'ivod'n'ь|, не обладает четкими характеристиками, позволяющими определить, что это|a|, а не |i| или |e|. Для письменной речи, которая, в отличие от устной, требует максимальной стандартизации, такое решение вопроса невозможно. Кроме того, оно противоречило бы и морфологической системе русского языка, как она представлена в сознании его носителей.

Основываясь на этих соображениях, Л. В. Щерба указывал на необходимость в письменной речи ориентироваться на полный тип произнесения. Следует заметить, что стихийно этому требованию следуют на практике всегда. Полный тип представлен таким произнесением, когда идентификация соответствующей фонемы не вызывает колебаний. Говорящие сознательно прибегают к полному типу тогда, когда они опасаются, что их могут не услышать или не понять. Часто это выражается в посложном произнесении, например: «Я сказал |v'e-čе-ram|, а не |n'e-čе-va|», или в специальном выделении слога, содержащего соответствующую фонему, например: «|paroda|, а не |paqoda|», и т. п.

Для того чтобы слово было опознано, какая-то часть его, т. е. часть фонем, оставляющих его облик, должна быть произнесена в полном типе. Тогда в силу избыточности языкового кода, обусловленной действием контекста, остальные фонемы, произнесенные в неполном типе, будут «реконструированы». В этом и состоит механизм опознания единиц спонтанной речи.

Вслед за Бодуэном де Куртенэ, в фонетическом типе письма принято различать три категории: алфавит, графику и орфографию.

Алфавит — это перечень букв и их звуковые значения. Внешний вид букв теоретически совершенно безразличен. Необходимо только, чтобы они хорошо различались между собой, были достаточно компактными и вместе с тем четкими, что особенно важно в наши дни, когда надо иметь в виду и пригодность шрифта для автоматического распознавания.

Обо всем этом можно говорить только в отношении младописьменных языков, так как переходить на новый вид письма в старописьменных языках далеко не всегда целесообразно. Что касается вновь создаваемых систем письма, то опыт показывает, что почти всегда алфавит заимствуется из какого-нибудь старописьменного языка. Теоретически для этого никаких препятствий нет, а практически это вполне удобно, особенно с полиграфической точки зрения, так как не требует перехода с

одного алфавита на другой. Разумеется, при заимствовании алфавита приходится дополнять его или сокращать или же вносить коррективы иного характера в соответствии с фонетической (точнее — фонологической) системой заимствующего языка.

Было бы ошибочным думать, что число букв в идеальном алфавите должно соответствовать числу фонем данного языка. Письмо — не фонематическая транскрипция, при которой требуется пофонемное отображение звукового облика слова. Письмо должно адекватно передавать этот облик, что вовсе не означает обязательности передачи каждой отдельной фонемы. Необходимо только, чтобы написание обеспечивало однозначное чтение языковых единиц, а для этого не требуется передавать на письме фонему за фонемой. Написания *соль* и *тюк*, например, достаточно точно отражают звуковой облик этих слов, т. е. всю цепочку составляющих этот облик фонем, хотя в первом из них три фонемы обозначены четырьмя буквами, а во втором палатализованный характер первой фонемы выражен буквой, обозначающей последующий гласный. Благодаря остроумному использованию букв русская азбука, состоящая из 33 букв, обеспечивает адекватную передачу 41 фонемы.

Таким образом, в рациональном алфавите может быть меньше букв, чем фонем, но не должно быть больше. Недопустимы дублирующие буквы, примером которых могут служить буквы *ѣ*, *і*, *ѳ* в русской азбуке до реформы 1918 г., имевшие то же значение, что буквы *е*, *и*, *ф*. В немецком алфавите лишними являются буква v, дублирующая f, и x, передающая сочетание фонем /ks/, которое может быть обозначено и действительно обозначается в ряде слов буквосочетанием chs.

Неправильное понимание сущности алфавита ведет к тому, что иногда в него включают и сочетания букв, служащие для обозначения одной фонемы. Так, в алфавите испанского языка мы наряду с буквой c, передающей переднеязычный щелевой согласный, и буквой h, не имеющей никакого звукового значения, находим их сочетание ch, служащее для обозначения аффрикаты /č/. Следуя этому примеру, нужно было бы включить в русский алфавит сочетания *бь, вь, гь, дь* и т. д., которые обозначают мягкие согласные, что было бы, конечно, совершенно неправильно. Во всех подобных случаях сочетания букв остаются сочетаниями и не становятся единицами алфавита, своего рода буквами. В теории письма они и имеют особое наименование — диграфы, триграфы и т. д. (см. ниже).

Буква — понятие обобщенное; печатная или рукописная, имеющая разный вид в зависимости от стиля шрифта и от почерка, буква остается той же при тождестве ее звукового значения. Что д, *д*, и *д* одна и та же буква, ни у кого не вызывает сомнения, она и называется во всех случаях «дэ». Теоретически особый случай представляют прописные, или заглавные,

буквы. Звуковое значение их всегда совпадает со значением строчных: *о* и *О*, например, всегда читаются одинаково. Однако прописные буквы имеют особую функцию, обозначая собственные имена и начало предложения. Учитывая фонетический тип русского письма, их следует признать разновидностями тех же букв, что и строчные, так как последние, в свою очередь, обозначают не собственные имена и не начало предложения. Рассмотренный случай позволяет сказать, что буквы выполняют не только собственно языковую, но и метаязыковую функцию. Такую же функцию выполняют также различные шрифты: жирный, светлый, курсив и т. п. Они служат для выделения отдельных частей текста, на которые пишущий хочет обратить особое внимание читающих.

Вторая категория письма — графика. **Графика** — это совокупность правил передачи средствами данного алфавита плана выражения языковых единиц, т. е. фонем и сочетаний фонем, входящих в звуковой облик этих единиц, безотносительно к плану содержания, т. е. без учета конкретных слов или морфем.

По правилам русской графики, например, фонема |r| во всех случаях передается буквой *р*, причем следующий за ней гласный, если он имеется, обозначается буквами *а, о, э, у, ы*. Фонема |u| в абсолютном начале слов и после твердых согласных обозначается буквой *у*, после мягких — буквой *ю*. Фонема |t| перед гласными и сонорными согласными передается буквой *т*, а перед пробелом (т. е. в конце слов) и перед глухими согласными — либо буквой *т*, либо буквой *д*. Это возможно благодаря тому, что по фонетическим законам современного русского языка оппозиция |t| и |d| в этих позициях нейтрализуется, и любое написание в этих случаях будет обязательно прочтено соответственно норме. Так, |trut|, например, будет прочтено одинаково, независимо от того, передано ли оно на письме как *трут* или *труд*. Выбор того или иного написания определяется тем, какое именно слово имеется в виду, но это уже сфера орфографии, о которой речь будет ниже, а не графики.

Как это видно из сказанного, одна и та же фонема может быть обозначена на письме разными буквами или сочетанием букв. Исходя из этого за минимальную единицу графики следует принять особую единицу; ее принято называть графемой. Г р а ф е м а — это класс букв и их сочетаний, служащих для передачи на письме соответствующей фонемы. Для обозначения, например, русской фонемы |t| служит графема *т/д*, складывающаяся из букв *т* и *д*. Каждая из таких букв, однозначно передающих фонему в данной позиции, представляет в ней соответствующую графему и называется поэтому а л л о г р а ф о м. Графема может быть представлена и одним аллографом (ср. *м, н, р*, обозначающие фонемы |m|, |n|, |r|). В качестве алло-

графа могут употребляться и сочетания букв — диграф (рус. *пь, сь, ль* и т. п., фр. ch|š|), триграф (нем. sch |š|).

Графема имеет полевую структуру: в центре находится основной, наиболее часто встречающийся аллограф; это обычно буква, употребленная в ее алфавитном значении, часто обнаруживаемом в ее названии. Например, основными аллографами графем, передающих русские фонемы |b|, |k|, |a|, |i|, будут буквы *б, к, а, и* (ср. их названия — *бэ, ка, а, и*).

Сложная структура графемы в старописьменных языках объясняется, с одной стороны, фонетическими изменениями, которые не сразу и не всегда находят отражение на письме (*ѣ* в дореформенном русском письме передавал особый гласный, слившийся впоследствии с |e|); с другой стороны, в некоторых системах письма существенное значение имеют заимствования. Так, в английском языке графема, обозначающая фонему |s|, имеет аллограф ps (psychic).

Графема как кодовая единица билатеральна, ее планом содержания является обозначаемая ею фонема. Различаясь в плане содержания, графемы могут иметь полностью или частично одни и те же аллографы, т. е. они могут пересекаться. В таких случаях в активном аспекте возникает омография, а в пассивном — полифония. Рассматривавшаяся выше графема для |t| совпадает по составу аллографов с графемой для |d|. В слове |addat'|, например, первое |d| передается через *т*, а второе — через *д* (*отдать*). Эти графемы, следовательно, омографичны. Разрешается омография графическим правилом, которое для пишущих гласит: «Фонема |d| обозначается буквой *д* во всех позициях, кроме позиции перед звонким согласным, где она может обозначаться и буквой *т*, например, *дом, куда, одна, отдать*. Фонема |t| обозначается буквой *т* во всех позициях, кроме позиции перед пробелом и перед глухим согласным, где она может обозначаться и буквой *д*, например, *коты, тот, ход, подтолкнуть*». Для читающих действует такое правило: «Буква *д* читается как |d| во всех позициях, кроме позиции перед пробелом и перед глухим согласным, где она читается как |t|. Буква *т* читается как |t| во всех позициях, кроме позиции перед звонким согласным, где она читается как |d|». Графемы в таких случаях различаются основными аллографами. В приведенных примерах графема, передающая фонему |d|, будет иметь основным аллографом букву *д*, а графема, передающая фонему |t|, — букву *т*; первую можно обозначить как *д/т*, вторую — *т/д*.

Широко распространено частичное пересечение графем, при котором совпадает только часть аллографов. Например, графема, обозначающая фонему |к|, имеет аллографы *к* и *г*, а графема, обозначающая фонему |x|, имеет аллографы *х* и *г* (*кусок, рог; мох, легко*).

Пересечение графем, даже и полное, как *т/д* и *д/т*, допустимо

и в рациональной графике при том условии, что омография и полифония снимаются графическими правилами.

Иное дело факультативная вариативность, которая заключается в том, что два знака имеют полностью совпадающую дистрибуцию. Это наблюдалось в русском дореформенном письме, где *ѣ, і, ѳ* во всех позициях обозначали те же фонемы, что *е, и, ф*.

Третьей категорией письма является **орфография**, которая представляет собой свод правил передачи на письме средствами графики данного языка плана выражения языковой единицы с учетом плана содержания. Последнее и отличает орфографию от графики. В соответствии с правилами орфографии производится выбор одного из возможных графических вариантов передачи на письме значимых языковых единиц — слов или морфем; например, из вариантов *сокол* и *сокал* для слова |sokal| выбирается первый, из вариантов для слова |ješčo| — *ещо, еще, есчо, есче, ешчо, ещче, ежчо, ежче, езчо, езче* — выбирается *еще*.

Правила такого выбора основываются на некоторых общих соображениях, которые называются принципами орфографии. Обычно орфографическая система конкретного языка строится с учетом не одного, а нескольких принципов, из которых какой-нибудь один является основным. Различаются следующие принципы орфографии: фонематический, морфематический, грамматический, дифференцирующий, традиционный, цитатный, транслитерационный и транскрипционный.

Написания, следующие фонематическому принципу, должны отражать звуковой облик слова посредством букв, употребляемых в их основном (алфавитном) значении. В соответствии с этим, следовательно, слово |rot|, например, независимо от его значения (*рот* или *род*) должно писаться с *т*. Пишущему остается только определить, начинать ли слово со строчной буквы или же с прописной. Это определяется правилом, имеющим не фонетическое, а метаязыковое значение. По правилам орфографии многих языков слово начинается с прописной буквы, если оно начинает предложение или же если оно начинает имя собственное. Фонематические написания широко используются в младописьменных языках, достаточно часто встречаются они и в старописьменных: ср. рус. *стол, дом, рука, лампа, кошка* и др.

Морфематический принцип требует, чтобы в написании морфемы во всех случаях отражался ее фонемный состав в сильных позициях. Живые чередования фонем, таким образом, на письме не отражаются. Так, во всех словоформах слова |got| *год* в корне по сильной позиции всегда пишется гласный *о* и в конце корня *д*. Точно так же в окончании тв. падежа ед. числа м. и ср. рода им. существительного |om| ~ |am| всегда пишется *о* — по сильной позиции (*глазом, зеркалом — сто-*

лом). Распространить применение морфематического принципа на исторические чередования означало бы замену фонетического типа письма идеографическим. Ср., например: *друг, дружба, дружей* вместо *друг, дружба, друзей.*

Грамматический принцип имеет чисто метаязыковую функцию. Согласно этому принципу в написании слова должно быть указание на соответствующую грамматическую форму. Наличие или отсутствие мягкого знака в глагольных формах типа *садится — садиться,* например, позволяет отличать 3-е лицо ед. числа наст. времени от неопределенного наклонения. Никакой фонетической реальности за подобными различиями нет.

То же можно сказать о дифференцирующем принципе, который часто и не отличают от грамматического. Смысл его заключается в том, что он позволяет графически различать омонимы. В некоторых языках он применяется довольно широко (например, в английском и французском), в русском он представлен в *компания — кампания, бал — балл.*

Как уже указывалось выше, письменная форма языковых единиц неизбежно отстает от устной. А происходит это оттого, что последняя изменяется спонтанно, а орфографическая норма создается сознательно, изменения в нее вносятся только тогда, когда противоречие между написанием и произношением становится очевидным. Поэтому в старописьменных языках всегда имеются такие написания, которые не оправданы современным состоянием языка. Встречаются среди них и такие, которые не отражают и старых форм, а возникли случайно. Поэтому наименование традиционный больше подходит к этому принципу, чем «исторический».

В лексике современных языков немалый удельный вес составляют заимствованные слова; они часто имеют особое написание, которое регулируется тремя принципами: цитатным, транслитерационным и транскрипционным.

Цитатный принцип применяется в трех случаях. Во-первых, когда в заимствующем языке не хватает нужных графических средств. Во-вторых, при сознательном стремлении отразить происхождение заимствования. В-третьих, как это имеет место в языках бывшего СССР, алфавит которых построен на основе русского, сохранение исконного написания слов русского происхождения мотивируется тем, что русский язык имеет широкое распространение как второй язык. При таких обстоятельствах было бы очень неудобно, если бы одно и то же слово писалось в одном тексте по-русски, а в другом по-иному.

Транслитерационный принцип заключается в побуквенной замене знаков заимствуемого слова знаками алфавита заимствующего языка, например: нем. *Berlin* — рус. *Берлин.* Ввиду несовпадения состава букв в разных алфавитах примене-

ние этого принципа не всегда возможно (ср. нем. *Marx* — рус. *Маркс*). Вследствие этого транслитерационный принцип неизбежно комбинируется с транскрипционным. Последний означает передачу фонемного состава заимствуемого слова, насколько это позволяют сделать графические средства. В имени Маркса это оказалось возможным, в имени Гегеля — нет, так как в русском алфавите нет знака, которым можно было бы передать немецкий согласный |h| (Hegel).

Первые три принципа основаны на закономерностях, присущих самому языку, а потому и мотивированных для носителя этого языка, позволяющих ему сознательно усваивать орфографию. В рациональной орфографии они должны быть абсолютно господствующими. Напротив, недопустим в такой орфографии традиционный принцип, совершенно немотивированный синхронным состоянием языка, превращающий изучение орфографии в зубрежку. Дифференцирующий, а также три принципа передачи заимствований, хотя и не мотивированы, но оправданы приведенными выше соображениями. При разумном использовании они допустимы.

Литература

Амирова Т. А. К истории и теории графемики. М., 1977.

Бодуэн де Куртенэ И. А. Об отношении русского письма к русскому языку. СПб., 1912.

Ветвицкий В. Г., Иванова В. Ф., Моисеев А. И. Современное русское письмо. М., 1974.

Волоцкая З. М., Молошная Т. Н., Николаева Т. М. Опыт описания русского языка в его письменной форме. М., 1964.

Гельб И. Е. Опыт изучения письма. М., 1982.

Гумбольдт, Вильгельм фон. Избранные труды по языкознанию. М., 1984.

Диринтер Д. Алфавит. М., 1967.

Зиндер Л. Р. Очерк общей теории письма. Л., 1987.

Истрин В. А. Возникновение и развитие письма. М., 1965.

Селезнева Л. Б. Современное русское письмо. Томск, 1981.

Соссюр, Фердинад де. Труды по языкознанию. М., 1977.

Фридрих И. История письма. М., 1979.

Щерба Л. В. Теория русского письма. Л., 1983.

ОРФОГРАФИЯ

Важнейшая роль в коммуникации принадлежит письму. Являясь определенной формой фиксации и хранения речи, письмо приобретает особое значение в эпоху научно-технической революции и всеобщей грамотности. Оно обеспечивает как непосредственную передачу научной информации, так и ее хранение и накопление. Всеобщая грамотность и высокая образованность общества повышают и широту распространения художественной литературы, для которой письменная форма закрепления является основной (чтение художественных произведений, газетных текстов и т. п. по радио или со сцены — это лишь эпизодическое использование устной формы).

Обращенность письменной формы речи к массовому потребителю требует рациональности устройства письма и его постоянного совершенствования, что является одним из аспектов прикладной лингвистики, поскольку последняя отражает социальные потребности общества.

Современное русское письмо, как и большинство современных систем письма, — буквенно-звуковое, точнее — фонемное: буквы как единицы письма обозначают фонемы в их временной последовательности. В системе письма выделяют три стороны (или аспекта, или фактора): алфавит, графику, орфографию (они были выделены, очерчены И. А. Бодуэном де Куртенэ). Наиболее изменяемым и подвижным фактором из этих трех является орфография. Именно орфография требует постоянного к ней внимания (со стороны языковедов) и охранительных мер.

Алфавит — это совокупность расположенных в определенном порядке букв в их связи с фонемами того или иного языка (они предназначены для обозначения фонем данного языка); графика — это правила отображения фонем и их сочетаний буквами данного алфавита без учета смысловых

единиц языка (морфем и слов), в которых они встречаются в потоке речи; орфография — это правила отображения фонем и их сочетаний на основе данного алфавита и правил графики, но уже с учетом морфемного состава слова, морфологических связей и отношений, а также (часто) и этимологии слова. Орфография представляет собой исторически сложившуюся систему единообразных написаний.

Развитие русского письма, его орфографии было постепенным. В его истории было только две реформы: петровская реформа азбуки (1708—1710) и реформа орфографии 1917—1918 гг. Петр I вместо церковной кириллицы, исключив из нее некоторые буквы, ввел новый гражданский шрифт, установил начертательное различие строчных и прописных букв (до этого использовались только прописные буквы), отменил обязательность постановки знаков ударения в каждом слове, что строго соблюдалось в печатных текстах, и др. Вторую реформу провело Советское правительство, хотя готовилась она в рамках Академии наук с начала XX в., 23 декабря 1917 г. Народный комиссариат просвещения издал декрет о введении нового правописания. 10 октября 1918 г. Советом народных комиссаров был издан специальный декрет, подтвердивший декрет Народного комиссариата просвещения. Второй декрет был вызван необходимостью расширить сферу применения новой орфографии, утвердить ее во всей советской печати. Реформа 1917—1918 гг. свела русское письмо с исторических рельсов. Она уничтожила уже совершенно чуждые духу русского языка церковнославянские окончания *-аго, -яго* у прилагательных, причастий и местоимений (*добраго, пятаго, котораго, синяго* и т. п.), заменив их русскими окончаниями *-ого, -его* (*доброго, пятого, которого, синего*). Было устранено разграничение форм *одни, они,* с одной стороны, и *однѣ, онѣ* — с другой (первые употреблялись для слов м. и ср. рода, вторые — для слов ж. рода). Оставлены лишь формы *одни, они* для всех трех родов. До реформы прилагательные, причастия и местоимения ж. и ср. рода в им. и вин. падежах мн. числа писали с окончаниями *-ыя, -ия* (*большия старыя липы*). Была уничтожена форма *ея* (*ея сестра*), была оставлена только форма *её* (*её сестра*). Уничтожена буква *ѣ* («ять») и другие дублетные буквы, еще остававшиеся в алфавите, и т. д.

Реформа 1917—1918 гг. была первой и пока единственной реформой русской орфографии (реформа Петра I была реформой азбуки). Упорядочение русской орфографии, проведенное (осуществленное) в 1956 г., не было реформой правописания, оно не затронуло его основ, хотя унификация колеблющихся написаний коснулась всех разделов правописания. Основным же результатом работы Правительственной орфографической комиссии (утвержденной в 1939 г.) было создание научно обоснованного свода «Правил русской орфографии и пунктуации»

(М., 1956), которые и до сих пор являются действующим законом современной орфографии.

Кроме названных реформ и упорядочения 1956 г. были еще два не проведенных в жизнь проекта реформ: 1) проект, подготовленный Сектором науки Наркомпроса РСФСР в 1930 г. (Проект Главнауки..., 1930); 2) проект, подготовленный Комиссией по усовершенствованию русской орфографии при Институте русского языка АН СССР и Отделении литературы и языка АН СССР в 1964 г. (Предложения по усовершенствованию русской орфографии, 1964). Первый из этих проектов декларировал равнение на малограмотных и неграмотных, в первую очередь (Проект Главнауки..., 1930, с. 10—11), приспосабливаясь к малограмотным написаниям; как научно не обоснованный, он был отвергнут.

Проект 1964 г. готовился в течение двух лет, но его подготовка сопровождалась большой и чрезвычайно интенсивной исследовательской работой. Его научная платформа отражала одно из научных направлений в теории орфографии (построенное на теории московской фонологической школы). Однако к обсуждению проекта не была подготовлена общественность, что показала развернувшаяся дискуссия. Особенно острой критике проект был подвергнут на страницах «Литературной газеты». Больше всего возражений вызвали первые четыре пункта проекта (изъять из алфавита *ъ*, заменив его на *ь*: *сьезд, обьявление* и т. п.; отменить употребление *ь* в грамматической функции: *ноч, доч, береч, не плач* и т. п.; писать *огурци, циган* и т. п.; писать по произношению *чорный, жоны, смягчонный, течот* и т. п.). В результате принятие предложений по усовершенствованию русской орфографии было отложено.

Но значение работы орфографических комиссий состоит не только в том, что они представляют предложения по усовершенствованию конкретных правил правописания. Пожалуй, еще большее значение этой работы состоит в том, что она способствует привлечению внимания к такому важнейшему участку языкового культурного фронта, как орфография. Возросшее внимание к орфографии способствует дальнейшей разработке ее теории и внедрению достижений в практику. Так, обсуждение «Предложений по усовершенствованию русской орфографии» 1964 г. и публикации подготовившей эти предложения комиссии внесли весомый вклад в науку.

Богатый теоретический материал содержат сборники, подготовленные этой комиссией: «Вопросы русской орфографии» (М., 1964), «О современной русской орфографии» (М., 1964), «Проблемы современного русского правописания» (М., 1964), «Орфография собственных имен» (М., 1965). Несомненно, был стимулирован работой комиссии и сборник «Нерешенные вопросы русского правописания» (М., 1974), хотя он вышел уже после того, как прекратилась работа комиссии. Но особое значе-

ние для науки имеет фундаментальная работа, выполненная по заданию комиссии, — «Обзор предложений по усовершенствованию русской орфографии (XVIII—XX вв.)» (М., 1965). В ней собраны все предложения по усовершенствованию русской орфографии, которые предлагались в течение 200 лет, с указанием первого их «автора» (а также всех последующих) и всесторонним анализом каждого предложения: какому орфографическому принципу отвечает то или иное предложение, какие у него плюсы и минусы. Самостоятельную ценность имеет библиография, насчитывающая более тысячи наименований.

Совершенствование русской орфографии продолжается, хотя и не в рамках реформы.

Большую роль в упорядочении русской орфографии играют словари.

Словарное богатство русского литературного языка очень велико. Ни в правилах орфографии, ни даже в специальных справочниках нельзя охватить всех случаев написаний, причем в языке есть значительное количество слов, орфография которых даже не определяется правилами. Сюда относятся как русские слова с непроверяемыми гласными и согласными, так и — особенно! — слова заимствованные.

Сразу же после выхода в свет «Правил русской орфографии и пунктуации» (1956) был создан большой академический «Орфографический словарь русского языка» (М., 1956) под редакцией С. И. Ожегова и А. Б. Шапиро. Подготовка орфографического словаря началась в учреждениях Академии наук еще до Великой Отечественной войны. Подготовительная работа велась под общим руководством академика С. П. Обнорского. В словаре 1956 г. охвачено 110 тыс. слов. Выход в свет «Орфографического словаря русского языка» должен был внести единообразие в русское правописание, а «упорядоченность орфографии, — по словам С. И. Ожегова, — расценивается как важный показатель культуры нации» (Вестн. АН СССР. 1957. № 1. С. 30).

К настоящему времени академический орфографический словарь («Орфографический словарь русского языка») выдержал 29 изданий (1-е изд. вышло в 1956 г., последнее, 29-е, — в 1991 г. Редакционная коллегия: В. В. Лопатин (отв. редактор), Б. З. Букчина, Л. П. Калакуцкая, Ю. Н. Караулов, С. М. Кузьмина). Первые четыре издания (1956, 1957, 1958, 1959) были стереотипными, 5-е издание (1963) — исправленное и дополненное (фактически это второе издание). В издании 1963 г. были учтены замечания и предложения, поступившие в сектор культуры речи Института русского языка от отдельных лиц и учреждений (в частности, от Комитета научно-технической терминологии АН СССР). Затем опять шли стереотипные издания (с 5-го по 12-е). Еще более значительному обновлению подверглось 13-е издание сло-

варя 1974 г. (фактически это третье издание словаря). В словарь было включено много новых слов, вошедших в русский язык в последние десятилетия. Внесены коррективы в написание отдельных слов (*хазары, бефстроганов, серосодержащий, соединительнотканный* и др.) в связи с появлением новых трудов, посвященных истории и функционированию отдельных слов, а также в связи с выходом новых, в частности, терминологических словарей и энциклопедических изданий. Последующие издания орфографического словаря (с 14-го по 28-е) — стереотипные. В 29-м издании изменены некоторые написания. Ликвидировано, например, исключение *розыскной*. Теперь — *разыскной*. Оставлен лишь один вариант из двух, а именно *приумножить* (ликвидирован *преумножить*), и др. Включено около 3000 слов; в основном это слова, вошедшие в русский язык в последние десятилетия. Таким образом, это издание — исправленное и дополненное. Фактически оно является четвертым изданием. 30, 31, 32-е издания — стереотипные.

В соответствии с «Правилами...» и академическим орфографическим словарем находится и школьный орфографический словарь. Первое издание этого очень важного словаря было осуществлено в 1934 г. Д. Н. Ушаковым. Позднее он стал словарем двух авторов — Д. Н. Ушакова и его ученика и последователя С. Е. Крючкова. В 1957 г. «Орфографический словарь для учащихся V—XI классов» Д. Н. Ушакова и С. Е. Крючкова вышел 13-м изданием. Оно было составлено на основе «Правил...». В 1972 г. вышло в свет 27-е издание этого словаря, переработанное и дополненное. Переработку словаря осуществил Л. А. Чешко. В словарь было включено значительное количество новых слов, были исключены устаревшие слова и формы. Последующие многочисленные переиздания этого словаря подготавливает Л. А. Чешко.

Большое теоретическое и практическое значение в плане усовершенствования русского правописания имеет создание дифференциальных орфографических словарей. Это — словарь «Слитно или раздельно?». 1-е издание — 1972 г. Составители — Б. З. Букчина, Л. П. Калакуцкая, Л. К. Чельцова (2-е изд. — 1976 г.). Начиная с 3-го издания составители — Б. З. Букчина, Л. П. Калакуцкая. Третье издание (1982, испр. и доп.) почти в два раза увеличило количество слов (82 000 по сравнению с 43 000 в первом издании). Последующие издания были стереотипными. Последнее издание — 7-е (М., 1990).

В 1984 г. был создан словарь-справочник «Прописная или строчная?». Автор — Д. Э. Розенталь. Автору удалось найти удачные принципы алфавитной подачи тех словосочетаний, у которых одно из слов пишется с прописной буквы. Для этого были использованы «ключевые» слова обобщенного типового значения. Словарь выдержал уже пять изданий. Последнее из-

дание, 5-е, исправленное (М., 1989), содержит около 8600 слов и словосочетаний.

Издан словарь «Слова с двойными согласными» Н. П. Колесникова (М., 1990).

В наш век, век невиданных темпов развития науки и ее отдельных отраслей, для упорядочения орфографического оформления многочисленной и разнообразной терминологической лексики должны были бы быть созданы отраслевые орфографические словари, но их почти нет. Есть «Орфографический морской словарь» Р. Э. Порецкой (1974) и «Орфографический словарь сложных прилагательных, используемых в медицинской литературе» (1978). Упорядочению орфографии способствуют и отраслевые терминологические словари, хотя они и не имеют прямой орфографической предназначенности.

Особого внимания в отношении унификации написаний требуют имена собственные и особенно географические названия. Различительные и адресные функции географических названий требуют их единообразного написания во всех официальных документах. В 1960 г. по постановлению Президиума АН СССР была организована специальная Комиссия по упорядочению написания и произношения собственных имен и географических названий ОЛЯ АН СССР. Сейчас вопросами стандартизации географических названий занимаются международные организации. При Организации Объединенных Наций создана постоянно действующая группа экспертов, занимающаяся проблемами стандартизации географических названий мира. Одна из основополагающих установок ее работы — опора на национальную стандартизацию. См. об этом подробнее в статье А. М. Комкова (входящего в качестве эксперта в эту группу) «Проблемы стандартизации (нормализации) географических названий в национальном и международном аспекте» (Ономастика и норма. М., 1976). Известной стандартизации географических названий способствуют такие, например, справочники, изданные в СССР, как «Словарь географических названий» М. С. Боднарского (2-е изд., испр. и доп. М., 1958), «*Словарь географических названий зарубежных стран*» под ред. М. Б. Волостновой (2-е изд., испр. М., 1988).

Однако наличие большого количества справочников и словарей, не подвластных никакому объединяющему авторитетному источнику, создает нежелательный орфографический разнобой. Наиболее яркий пример — это разночтения в «Орфографическом словаре русского языка» и дифференциальном словаре «Слитно или раздельно?». При выборе слитного или дефисного оформления сложных прилагательных авторы словаря «Слитно или раздельно?» опираются на выделенный и сформулированный ими словообразовательно-грамматический принцип: наличие суффикса в первой части сложных прилагательных определяет дефисное написание (*автомобильно-дорожный*,

угольно-графитовый), а его отсутствие определяет слитное написание (*автодорожный, углеграфитовый*). Стремление именно так оформлять сложные прилагательные было замечено орфографистами в печатной практике. Часто оформление по этому принципу шло вразрез с «Правилами русской орфографии и пунктуации» 1956 г. Так, по правилам нужно писать *сравнительноанатомический* (так как оно «раскладывается» на подчинительное словосочетание *сравнительная анатомия*), но печать упорно давала *сравнительно-анатомический*. И таких случаев было много. Стремление пишущих оформлять сложные прилагательные на основе словообразовательной структуры, а не на основе синтаксических связей и отношений в сконструированном словосочетании (которые не всегда, кстати, можно и сконструировать), было названо «орфографической тенденцией» (Букчина Б. З., Калакуцкая Л. П. Сложные слова. М., 1974). Правописание на основе этого принципа и проведено авторами в словаре «Слитно или раздельно?». Но написания в этом словаре, выдержанные в «ключе» орфографической тенденции, как уже сказано, иногда расходятся с рекомендациями «Орфографического словаря русского языка». В «Орфографическом словаре...»: *западноевропейский, восточнославянский*; в словаре «Слитно или раздельно?»: *западно-европейский, восточно-славянский*. Возникшая вариантность нежелательна. Языковой «истине» больше отвечают последние написания. В последнем, 29-м издании «Орфографического словаря...» (1991), сделана известная уступка словарю «Слитно или раздельно?» (предложено, например, писать через дефис *народно-хозяйственный* и др.), но уступка эта проведена непоследовательно: слова одинаковой структуры пишутся поразному, то в соответствии со словарем «Слитно или раздельно?», то не в соответствии, что затрудняет пользование 29-м изданием «Орфографического словаря».

Но не только с неизбежностью возникающий орфографический разнобой вызывает необходимость постоянного усовершенствования орфографии. В русском письме (и в орфографии, и в пунктуации), как и в письме всех народов вообще, возникают новые правописные ситуации. Появляются новые слова, возникают новые синтаксические конструкции и т. п. Все это нуждается в оформлении. И тогда рождаются новые приемы оформления слов и конструкций. Приемы оформления новых правописных ситуаций складываются то на основании подравнивания к существующим моделям, то на основании счастливых находок и открытий (на основе языкового сознания пишущих). Пунктуация, как известно, вообще складывалась под пером писателей и в искусстве типографских мастеров. Правила пунктуации выводились из практики. И сейчас практика нашла много приемов и в оформлении слов, и еще больше — в оформлении синтаксических конструкций.

Такого нового, не охваченного «Правилами...» 1956 г., сейчас накопилось много, и перед языковедами стоит задача обобщить и осмыслить новый орфографический материал.

Нельзя сказать, что такая работа не ведется. Исследования постепенно накапливаются. Новые правописные тенденции отражают различные справочники (особенно для работников печати). Но вся эта работа нуждается в целенаправленном упорядочении, поскольку не только в пособиях, написанных отдельными авторами, но даже и в словарях, как видим, рекомендации не всегда однозначны.

Необходимо все это «уединообразить». Такая координационная работа подвластна лишь компетенции Института русского языка РАН. В 1990 г. при этом институте образована новая Орфографическая комиссия.

Литература

Аванесов Р. И. Заметки по теории русской орфографии // Восточнославянское и русское языкознание. М., 1978.

Амирова Т. А. К истории и теории графемики. М., 1977.— Библиогр.

Богоявленский Д. Н. Психология усвоения орфографии. 2-е изд. М., 1966.

Бодуэн де Куртенэ И. А. Об отношении русского письма к русскому языку. СПб., 1912.

Букчина Б. З. Вопросы орфографической унификации печатных изданий. М., 1988. — Библиогр.

Вахек Й. К проблеме письменного языка // Пражский лингвистический кружок. М., 1967.

Ветвицкий В. Г., Иванова В. Ф., Моисеев А. И. Современное русское письмо: Пособие для учителей. М., 1974.— Библиогр.

Волков А. А. Грамматология: семиотика письменной речи. М., 1982.

Волоцкая З. М., Молошная Т. Н., Николаева Т. М. Опыт описания русского языка в его письменной форме. М., 1964.

Вопросы русской орфографии. М., 1964.

Гак В. Г. Орфография в свете структурного анализа // Проблемы структурной лингвистики. М., 1962.

Гвоздев А. Н. Основы русской орфографии. 4-е изд. М.. 1954.— *То же:* Гвоздев А. Н. Избранные работы по орфографии и фонетике. М., 1963. С. 213—257.

Гельб И. Е. Опыт изучения письма (основы грамматологии). М., 1982.

Глисон Г. Введение в дескриптивную лингвистику. М., 1959. Гл. XXI.

Зиндер Л. Р. 1) Общая фонетика. 2-е изд. М., 1979. Гл. VIII; 2) Очерк общей теории письма. Л., 1987.

Иванова В. Ф. 1) Трудные вопросы орфографии. 2-е изд. М., 1982. — Библиогр.; 2) Современная русская орфография. М., 1991. — Библиогр.

Иванова В. Ф., Осипов Б. И. Принципы орфографии и их педагогическое значение // Рус. язык в школе. 1991. № 5.

Касаткин Л. Л. Графика и орфография // Современный русский литературный язык / Под ред. П. А. Леканта. 2-е изд. М., 1988.

Клибанова Л. А. Лингводидактическое описание орфографии русского языка как иностранного (к постановке проблемы). Кишинев, 1989. — Библиогр.

Кузьмина С. М. Теория русской орфографии. Орфография в ее отношении к фонетике и фонологии. М., 1981.— Библиогр.

Маслов Ю. С. Заметки по теории графики // Исследования по языку и литературе: Памяти В. М. Жирмунского. М., 1973.

Мейеров В. Ф. Типология буквенных орфограмм. Иркутск, 1988. — Библиогр.

Моисеев А. И. 1) Звуки и буквы, буквы и цифры... М., 1987; 2) 1000-летие русского письма // Рус. язык за рубежом. 1988. № 6.

Нерешенные вопросы русского правописания. М., 1974.

Обзор предложений по усовершенствованию русской орфографии (XVIII—XX вв.). М., 1965. — Библиогр.

Опыт совершенствования алфавитов и орфографий народов СССР. М., 1982.

Орфография собственных имен. М., 1965.

Осипов Б. И. История русской орфографии и пунктуации. Новосибирск, 1992. — Библиогр.

О современной русской орфографии. М., 1964.

Панов М. В. Теория письма. Графика. Орфография // Современный русский язык / Под ред. В. А. Белошапковой. 2-е изд. М., 1989. — Библиогр.

Предложения по усовершенствованию русской орфографии / Отдние лит. и яз. АН СССР. Ин-т рус. языка. Комиссия по усовершенствованию русской орфографии. М., 1964.

Проблемы современного русского правописания. М., 1964.

Проект Главнауки о новом правописании: Сборник. Статьи о принципах реформы орфографии и пунктуации. Постановления комиссии Главнауки. Комментарии. Образцы текстов по новому правописанию. Библиография / С предисл. редакции «За коммунистическое просвещение»; Сост. Н. Кременский и В. Мамонтов; Под ред. Г. Данилова и Г. Костенко. М., 1930.

Проект реформы правописания / Под ред. К. Бархина, Г. Костенко и И. Устинова. Сектор науки Наркомпроса РСФСР. Издание официальное. М.; Л., 1930.

Розенталь Д. Э. Вопросы русского правописания. 3-е изд. М., 1970.

Селезнева Л. Б. 1) Современное русское письмо (системный анализ). Томск, 1981; 2) Укрощаем орфограмму (алгоритмизированное обучение). Волгоград, 1993. — Библиогр.

Солнцев В. М. Язык и письмо: Предисловие // Амирова Т. А. Функциональная взаимосвязь письменного и звукового языка. М., 1985.

Соссюр Ф. де. Курс общей лингвистики. М., 1933.

Шапиро А. Б. Русское правописание. 2-е изд. М., 1961.

Щерба Л. В. 1) Основные принципы орфографии и их социальное значение // Избранные работы по русскому языку. М., 1957; 2) Теория русского письма. Л., 1983.

3 Заказ № 171

ОРФОЭПИЯ И ВОПРОСЫ КУЛЬТУРЫ РЕЧИ

Для исследования литературно-языковой нормы вообще и произносительной нормы в частности необходимо изучение не только внутренних системных, но и социолингвистических факторов, кроме того, определенных психологических законов, которые управляют речевой деятельностью людей. До сих пор остается нерешенным вопрос, существует ли норма в самом языке или она вносится в язык извне, противопоставляется ли норма кодификации или норма не отличается от кодификации.

Бо́льшая часть лингвистов считает норму лингвистической категорией, считает, что норма существует в самом языке. Такое понимание ориентировано в общем виде на идею противопоставления языка и речи, высказанную Ф. де Соссюром. Не вникая в некоторые различия в трактовке этого противопоставления разными авторами и не обращая внимания на некоторые терминологические различия, можно сказать, что все они рассматривают язык как систему условных средств, а речь — как форму существования системы. Именно это противопоставление языка и речи привело к противопоставлению системы и нормы, интуитивно ощущаемому многими русскими и зарубежными лингвистами (Соссюр, 1939). Наиболее последовательно разграничил систему и норму уругвайский лингвист Э. Косериу (1963). Стремлением преодолеть некоторые противоречия Э. Косериу были вызваны работы В. А. Ицковича (1968, 1970), Л. И. Скворцова (1970), Ю. А. Бельчикова (1965), Н. Н. Семенюк (1967) и др., однако все эти исследователи соглашались с Э. Косериу в главном: норма существует в самом языке, она определенным образом связана с системой, определяется системой. Именно эта основная идея и является важным положением данного раздела. Система языка — это система возможностей, существующих в конкретном языке, система моделей, полностью не реализованная.

С этой точки зрения кажется совершенно справедливым утверждение Л. А. Булаховского, что «никогда, ни в какую

эпоху не были и не могли быть использованы до конца все формальные возможности, представляемые языком» (Булаховский, 1952, с. 130).

Языковая норма — совокупность наиболее устойчивых, *традиционных элементов системы*. Каждой системе свойственна омонимичность средств, ведущая к вариативности. В норме происходит отбор того, что имеется в системе. Если признавать, что норма присуща языку, что она является категорией лингвистической, то она должна быть у́же системы; норма не может выходить за пределы системы. Явлений, противоречащих системе, в норме быть не может; если же они возникают, то только после соответствующей перестройки системы, но и в этом случае новой системе явления эти уже противоречить не будут.

Таким образом, изменения нормы, которые мы наблюдаем, разрешены системой. По-видимому, *взаимоотношения произносительных лексических, морфологических, синтаксических и других норм и соответствующих систем будут несколько иными*, но во всех случаях норма не может быть шире системы.

Кодификация нормы — отражение нормы в учебниках, словарях, справочниках. Основной принцип кодификации — адекватность современной норме, — как правило, нарушается, так как кодификация отстает от нормы, фиксирует то, что существует в течение определенного времени.

Таким образом, под нормой следует понимать совокупность явлений, разрешенных системой языка, отобранных и закрепленных в речи носителей языка и являющихся обязательными для носителей языка.

Интересным с точки зрения обсуждаемой проблемы сущности нормы и различения нормы и кодификации кажется предложение Шиндлера, автора экспериментально-фонетического исследования немецкого литературного произношения, различать два *типа норм*: 1) *прескриптивную* (предписываемую) (т. е., по существу, кодификацию) и 2) дескриптивную (реализуемую) (т. е. *норму как внутриязыковую категорию*) (Schindler, 1974). Удачными представляются и термины, выбранные автором.

Признание нормы лингвистической категорией, признание объективной природы нормы не исключает сознательного вмешательства, нормализаторской работы, однако успешной такая работа станет в тех случаях, когда она будет соответствовать внутренним законам развития языка.

Вариативность — основная черта нормы, вариативность обязательно присуща языку, так как носители языка неминуемо реализуют различные способы произнесения тех или иных звуков, различные способы образования слов, связи слов в предложениях и т. п.

Проблема нормы возникает в тех случаях, когда в языке имеется не одна, а несколько возможностей осознания одной и той же языковой сущности. Можно ли обе эти возможности считать нормативными или только одну?

Изменение нормы обычно происходит через стадию существования двух способов выражения, в течение какого-то времени признаваемых одинаково правильными; вряд ли можно говорить о единственно нормативном способе выражения. Вариантность нормы (в отличие от вариативности, обязательной черты нормы) может быть как бы двух типов. Вариантность первого типа — существование двух или нескольких вариантов, признаваемых одинаково правильными. При вариантности второго типа они уже выстроились в определенный ряд: первый член этого ряда нормативен, другие употребляются все реже, становятся архаичными. Объяснить, почему укрепляется именно тот или иной вариант, оказывается порой очень сложно, а иногда и невозможно. В ряде случаев причины могут иметь и экстралингвистический характер.

При рассмотрении одного из частных вопросов общей проблемы нормы — произносительной нормы — целесообразно различать два аспекта произносительной нормы — орфоэпию и орфофонию. Отношения между этими аспектами могут рассматриваться по-разному в связи с различным пониманием фонемы. Исходя из щербовского понимания фонемы можно сказать, что орфоэпия занимается выявлением фонемного состава слова (например, в слове *шоссе* — фонема |а|, а не |о|, т. е. |šasse|, а не |šosse|), орфофония занимается вопросами качества произнесения отдельных звуков, тем, как должен произноситься тот или иной звук (например, в слове *шоссе* |е| — открытый гласный |ε|, |š| — твердый, веляризованный).

В самом общем виде можно сказать, что орфоэпия — это правила, определяющие нормативный фонемный состав слов, а орфофония — это правила произношения вариантов фонем. Так, вопрос о том, следует ли употреблять в слове *тихий* твердое |х| или мягкое |х'|, регулируется правилами орфоэпии, а требование произносить согласный |j| в конце слова как сонорный, а не глухой, относится к правилам орфофонии.

Орфоэпия для устной речи играет роль, подобную той, которую играет орфография для письменной речи (орфография определяет буквенный состав слова на письме, орфоэпия — фонемный состав). Сознательной кодификации подвергается именно орфоэпия, так как фонемный состав слов хорошо осознается носителями языка.

Орфофонические различия носителями языка часто не осознаются. Вместе с тем диалектные и просторечные явления чаще всего проявляются не в нарушении фонемного состава слова, а в изменении фонетических характеристик фонем, в ис-

кажении звуковой реализации фонем (например, недостаточная редукция безударных гласных, недостаточная мягкость |č'|).

Проведение последовательного различения орфоэпии и орфофонии при установлении произносительной нормы, при исследовании взаимоотношений нормы и вариантов представляется очень важным.

Ограничиться же при исследовании произносительной нормы только орфоэпией нельзя: так, произношение фрикативного |γ| вместо взрывного |g|, например в слове *голова*, недопустимо, хотя ошибка эта не орфоэпическая (в русском языке нет фонемы |γ|), а орфофоническая.

Проведение последовательного разграничения орфоэпии и орфофонии приводит к различению орфоэпической и орфофонической вариантности, причем, как уже говорилось, вариантности двух типов. Так, примером орфоэпической вариантности первого типа (существование двух однозначных способов выражения одной и той же языковой сущности) является произношение мягкого |s'| и твердого |s| в сочетании |st'| (например, слово *степь* может произноситься как |st'ep'| и как |s't'ep'|). Примером орфоэпической вариантности второго типа (выстраивание вариантов в определенный ряд) может быть произношение мягкого |s'| и твердого |s| в возвратных частицах глаголов: вариант с мягким |s'| |barols'a| является сейчас ведущим вариантом нормы. Пример орфофонической вариантности первого типа — реализации фонемы |č'|, имеющие некоторые различия в степени палатализации; пример орфофонической вариантности второго типа — произнесение конечных сонантов как сонорных или шумных |boj|, |bor|; однако ведущим является произнесение сонорных в конце слов.

В основном трудности исследования орфоэпических и орфофонических особенностей произносительной нормы связаны со сложностями фиксации произношения. Морфологические, синтаксические словообразовательные нормы в большей степени фиксируются письменной формой языка, облегчая тем самым возможность их изучения в плане диахронии и синхронии. Произношение же фиксируется с помощью фонетической транскрипции или словесного описания характера звуков и их сочетаний. В последние годы исследователи произношения получили возможность широкого использования магнитных записей.

Однако наибольшие трудности исследования произношения связаны с решением двух основных вопросов: как выбрать объекты исследования и какой методикой пользоваться.

Д. Н. Ушаков, много и плодотворно занимавшийся вопросами произносительной нормы, сам представлявший образец нормативного произношения, писал: «Понятие правильности условно: правильно то, что общепринято» (Ушаков, 1928, с. 5). Од-

нако, насколько можно судить по его работам, говоря об общепринятости, он имел в виду достаточно узкий круг носителей языка. Е. Ф. Петрищева прямо говорит о том, что норма должна строиться с учетом реальной языковой нормы образованных людей (Петрищева, 1967). Л. В. Щерба предлагал искать норму в произведениях классиков литературы — Горького, Чехова, Короленко, Тургенева, Гончарова и др. Представители пражского лингвистического кружка считали, что норму можно устанавливать исходя из языковой практики хороших авторов и речи интеллигентных слоев (Щерба, 1939).

Следует иметь в виду, что понятия нормированного литературного языка и языка художественной литературы не всегда совпадают. Стремясь к реалистическому изображению жизни, писатели включают в свои произведения диалектизмы, просторечия и т. п. Именно поэтому следует, по-видимому, разграничивать собственно авторскую речь и речь героев и не отождествлять нормативный литературный язык и язык хороших писателей.

Вопрос о выборе объектов исследования до сих пор остается нерешенным. По-видимому, следует признать, что объектами исследования должны быть образованные люди, в речи которых отсутствуют просторечные, диалектные, региональные особенности, какие бы то ни было яркие черты, указывающие на принадлежность к узкоограниченной социальной группе. В последние годы наметилась тенденция к определению нормы как некой отрицательной категории: нормативной можно считать речь, по которой ничего нельзя сказать ни о происхождении носителя языка, ни о его социальном положении.

При решении вопросов, связанных с произносительной нормой, методика исследования приобретает особое значение, так как результаты, полученные разными исследователями, занимающимися изучением произносительной нормы, часто оказываются различными из-за несовершенства методики исследования. Результаты анализа должны оцениваться в зависимости от способов их получения.

Наиболее распространенным методом исследования произношения является метод анкет, вопросников. Испытуемые должны дать ответы на вопросы, поставленные исследователем в анкете-вопроснике, в письменной форме. Такой метод исследования является широко распространенным прежде всего потому, что дает возможность получить большой статистический материал. Так, авторы социально-лингвистического исследования современного русского литературного языка получили ответы на вопросник по словообразованию от 3000 человек, а на вопросник по морфологии — от 4300 (Русский язык, 1974, с. 31).

Роль вопросников в лингвистических исследованиях акцентной, морфологической, словообразовательной вариантности велика. Однако применение этого метода для исследования про-

изношения вызывает определенные опасения. Носитель языка, заполняющий вопросник, должен решить, какое *д* произносит он в слове *дверь*: твердое или мягкое; сравнивая слова *истечь* и *испортить*, определить, в каком из них *с* произносится мягче, и т. д. По ответам информантов исследователь делает выводы о распространении того или иного произношения. Материалом для выводов о произносительных нормах являются свидетельства самих носителей языка. Вместе с тем полагаться на возможность носителя языка правильно определить, что и как он произносит, довольно трудно. Даже в тех случаях, когда информанты обладают фонетическим слухом, стремятся дать объективные ответы, обладают определенными знаниями, необходимыми для ответов, соответствующих действительности, установить нормативное произношение (не решая в данном случае вопроса о том, обязательно ли нормативное произношение — наиболее распространенное) с помощью такой методики по-видимому, не всегда возможно. В лучшем случае можно получить ответ на вопрос об осознании нормы носителями языка. Ответы информантов, получаемые в вопросниках, часто не отражают реального произношения, о чем свидетельствует сравнение данных вопросников с данными, полученнными при анализе магнитных записей фраз, составленных из слов, вошедших в вопросник. Так, например, по вопроснику частота мягкого *з* (слово *зверей*) 18,5%, по магнитной записи — 72%; мягкого *с* (слово *после*) по вопроснику — 27,4%, по записи — 62% (Баринова и др., 1971). Расхождения эти оказываются наиболее значительными в тех случаях, когда возможны колебания в произношении; достаточно точная самооценка наблюдается в тех случаях, когда исследуемый вариант, по существу, является единственным.

Все рассмотренные недостатки метода вопросников показывают необходимость выработки и применения иной методики исследования произносительной нормы. Необходимо получить не данные самоанализа произношения испытуемых, а реальные фонетические данные.

При исследовании произносительной нормы нужно стремиться к получению магнитных записей большого числа дикторов, однако следует иметь в виду, что в любом случае это будет лишь небольшая часть носителей языка (даже 3000 ответов, полученных методом вопросников, могут оказаться недостаточной выборкой). Запись подготовленного материала в произнесении подобранных дикторов на магнитофон дает возможность прослушать материал нужное число раз, подготовить в случае необходимости опыты по восприятию, производить инструментальный анализ материала.

Первым этапом анализа должен быть слуховой анализ записанного материала исследователем. Однако объективность полученных исследователем данных должна быть проверена с

помощью опытов по восприятию того же материала группой рядовых носителей языка (аудиторов), не посвященных в существо работы, так как исследователи могут подсознательно «слышать» то, что соответствует имеющейся у них гипотезе. Следует при этом заметить, что восприятие аудиторов тоже в определенной мере субъективно, но при достаточной статистике через него можно выявить объективные свойства системы или нормы. Однако определенные виды опытов по восприятию требуют отбора аудиторов и даже определенной их подготовки. Имеются в виду, например, опыты по идентификации изолированных звуков, выделенных из слов с помощью сепаратора. Предварительная подготовка заключается в прослушивании записанных на магнитофон таблиц бессмысленных слогов, чтобы приучить носителей языка к внимательному восприятию именно звуковой последовательности. Для участия в опытах отбираются те, кто при прослушивании слоговых таблиц дал не менее 95% правильных ответов.

В разных исследованиях опыты по восприятию могут иметь различный удельный вес, но в ряде случаев нельзя ограничиться только ими. Исследование произношения требует применения целого комплекса экспериментально-фонетических методов.

Кроме того, в связи с тем, что норма вообще и произносительная норма в частности в определенной мере понятие оценочное, целесообразно проведение опытов «вкусовой» оценки произносительных вариантов носителями языка: предъявляется два орфоэпических или орфофонических варианта, и от аудиторов требуется определить, какой из предъявленных вариантов кажется им правильнее, красивее, эстетичнее (объяснять причину того или иного решения не требуется). О возможности получения данных путем оценочного чувства правильности — неправильности носителей языка писал Л. В. Щерба (1974, с. 33).

Применение всего комплекса экспериментально-фонетических методов исследования произношения (слуховой анализ исследователей, различные типы опытов по восприятию носителей языка, осциллографический и спектральный анализ) будет показано далее на примере исследования «ленинградского» варианта произносительной нормы.

В течение нескольких десятилетий существовало два равноправных варианта произносительной нормы: так называемый московский и петербургский (ленинградский). Московский вариант произносительной нормы сложился раньше ленинградского. В основе петербургского произношения, по мнению В. И. Чернышева (1912, с. 12—13), лежит московский говор, подвергшийся воздействию окружающих Петербург северновеликорусских и средневеликорусских говоров. Рассмотрение произносительной нормы требует подробного, исчерпывающего описания двух равноправных его вариантов. Московский ва-

риант, который многие исследователи считали основным, довольно подробно и систематично описан в литературе. Специального описания ленинградского варианта в литературе по русистике нет, однако имеется целый круг явлений, классифицируемых как ленинградские особенности произношения (Аванесов, 1972). В фонетической литературе можно отметить 50 орфоэпических и орфофонических особенностей ленинградского произношения в области произнесения отдельных гласных и согласных, сочетаний согласных и отдельных слов; 39 из них являются особенностями ленинградского варианта нормы, остальные — ленинградского просторечия; 30 особенностей орфоэпических (8 — в области гласных, 22 — в области согласных) и 9 — орфофонических. Роль этих особенностей в речи ленинградцев разная. 40% особенностей связаны с произнесением больших групп слов: произношение безударного |e| на месте орфографических *я* и *е* после мягких согласных в предударных и заударных, открытых и закрытых слогах (т. е. |v'edu|, |płačet|, |pol'eⁱ| вместо московского |v'idu|, |plač'it|, |pol'ь|); произношение |i| в безударных флексиях полных прилагательных м. рода им. падежа после заднеязычных *г, к, х* (|gromk'ij| вместо |gromkъj|); произношение |a| в заударной флексии 3-го лица мн. числа глаголов 2-го спряжения (т. е. |xod'at| вместо |xod'ut|); отсутствие смягчения первого согласного в сочетаниях двух согласных при втором мягком (|kapl'eⁱt| вместо |kap'l'eⁱt|); отвердение губных в конце слов и перед |j| (произнесение слов *семь, кровь, подъем* как |s'em|, |krof|, |padjom|); произнесение мягкого |s'| в возвратных частицах глаголов (|uč'ils'a| вместо |uč'ilsa|); произношение долгого |n:| в ряде суффиксов с одним *н* и т. п. 40% особенностей связаны с произнесением ограниченного числа слов: невозможность утраты лабиализации |u| в заударном закрытом слоге (|čel'us't'| в отличие от |č'el'ьs't'|); произношение открытого |ɛ| на месте орфографического *е* после твердых губных (например, произношение слова *цвет* как |cvet| вместо |cv'et|); влияние мягких |l'|, |n'| на соседние согласные; произношение мягкого |r'| в корневых сочетаниях типа |sv'er'xu|, |s'er'p|; произнесение сочетания *щн* как |šn|; произношение сочетаний *чн* на стыке корня и суффикса как |č'n| вместо |šn| и т. п. 20% особенностей связаны с произнесением отдельных слов: произнесение слов *что, кто, никто* как |č'to|, |kto|, |n'ikto|); произнесение слова *дождя* как |dažd'a| вместо |daž':a|; слова *дождь* как |došt'| вместо |doš':|; произнесение сочетания |kk| вместо |xk| в слове *легка*; произнесение твердого |s| в слове *отсюда*.

Из 9 орфофонических особенностей 5 — в области гласных (бóльшая длительность ударных гласных, более закрытый характер ударного |a|, более открытый характер ударного |o|, бóльшая закрытость ударного |e|, отсутствие назализации

ударных гласных в абсолютном конце слов, отмечаемой в московском варианте) и 4 — в области согласных (произнесение сочетаний *сч, зч, щ* как |šč'| в отличие от |š':|; слабая палатизация |č'|, некоторое смягчение аффрикаты |c| перед в заимствованных словах, например |l'ekc'iju|; произнесение мягких |š'|, |ž'| в положении перед и после |l'|, |n'|).

Было проведено специальное исследование с целью выяснить, насколько перечисленные особенности характерны для речи ленинградцев.[1] В качестве объектов исследования было взято 150 человек без специального выбора. Все дикторы — жители Ленинграда, но по образованию, возрасту, диалектной базе и индивидуальным фонетическим особенностям они составили различные группы.

Из общего числа испытуемых 13 человек представляют коренных ленинградцев в 4-м поколении; исследование их произношения дает возможность получить данные об особенностях собственно «петербургско-ленинградского» произношения. Другая группа испытуемых состоит из лиц, живших в Ленинграде 16 и более лет, но до этого живших в различных городах страны и имевших разнообразную диалектную базу. Третья группа испытуемых проживала в Ленинграде менее 16 лет, но больше 5. С точки зрения возраста дикторы распределились следующим образом: 44% дикторов в возрасте от 18 до 25 лет, 15% — от 26 до 30 лет, 26% — от 31 до 35 лет и 16% — от 36 до 60 лет.

Основная масса дикторов в возрасте до 35 лет имела среднее общее, среднее специальное образование. Дикторы старше 35 лет имели высшее образование. Были предприняты попытки записать дикторов старше 35 лет и не имеющих высшего образования. Однако представители этой возрастной группы не имели, как правило, и полного среднего образования. Эта случайная выборка в определенной мере отражала реальный демографический состав жителей Ленинграда.

Испытуемым предлагалось прочитать список слов и специально составленный текст. При составлении экспериментального материала учитывались все потенциальные особенности ленинградского произношения, и поэтому в его состав были включены такие слова и сочетания, в которых эти особенности могли проявиться.

Чтение каждого диктора записывалось на магнитную ленту для дальнейшего анализа. Кроме чтения была получена и так называемая спонтанная речь, включавшая в себя сообщение сведений о себе и пересказ известной сказки. Таким образом, условия получения материала обеспечивают и реализа-

---

[1] С середины 60-х годов до настоящего времени проводится экспериментально-фонетическое исследование произносительных норм. Специальное исследование, о котором идет речь, закончено к 1990 г.

цию специфического вида литературной речи — чтения, и наиболее обычного вида — спонтанной речи.

Весь материал, записанный на магнитную ленту, был подвергнут анализу тремя различными способами: слуховому анализу, анализу восприятия различных звуковых элементов носителями языка, инструментальному (осциллографическому и спектральному) анализу[2].

Слуховой анализ материала, записанного исследователями, показал, что из 39 орфоэпических и орфофонических особенностей реально в речи исследованных дикторов встретилось лишь 19, т. е. около половины: 5 в области вокализма и 14 в области консонантизма и консонантных сочетаний (причем все особенности орфоэпические, за исключением одной орфофонической). Четыре орфоэпические особенности ленинградского произношения (две в области безударных гласных и две в области произнесения согласных) свойственны теперь общелитературной норме (они встретились в речи всех дикторов). Речь идет о произношении |i| в безударных флексиях полных прилагательных м. рода после *г, к, х,* произношении |a| в заударных флексиях 3-го лица мн. числа глаголов 2-го спряжения, произношении |s'| в возвратных частицах и ограничении ассимилятивного смягчения согласных в сочетаниях двух согласных при втором мягком. Бо́льшая же часть конкретных проявлений старых ленинградских черт не была последовательно представлена ни во всей группе испытуемых, ни в речи отдельных испытуемых. Оказалось также, что в спонтанной речи и при чтении эти особенности проявляются по-разному: чаще в чтении, чем в спонтанной речи.

В результате статистической обработки материала выяснилось, что наиболее распространенной чертой ленинградского произношения является «еканье», однако эта особенность сохранилась лишь в произношении части дикторов не обязательно во всех случаях, где она могла проявиться. Средний процент «еканья» по всем позициям в чтении — 20%, в спонтанной речи — 11%. Чаще оно сохраняется в первом предударном слоге и в абсолютном конце. Однако нет ни одного испытуемого, последовательно реализующего «еканье».

Следует отметить, что слуховая оценка качества гласного, классификация его как |e, e'| или |i| часто вызывала определенные затруднения и сомнения в правильности даваемой оценки. Следовало проверить, совпадает ли восприятие исследователей с восприятием носителей языка, не посвященных в существо работы. Было проведено 4 серии опытов. I — опознание гласных без ограничения возможности выбора; II — опознание безударных гласных в слоге согласный+гласный;

[2] Осциллограф Н-102, вибраторы типа II (собственная частота колебаний 10 000 Гц) и типа VIII (1200 Гц), скорость съемки 250 мм/с, спектрограф типа «видимая речь», скорость съемки 120 мм/с.

III — прослушивание с выбором из двух возможных звуков (аудиторы были предупреждены, что звучит только |e| и |i|); IV — оценка сходства и различия звуков.

В результате этих опытов было установлено, что только 20% звуков, опознанных исследователями как |e|, аудиторами было опознано так же. Это объясняется, по-видимому, тем, что ситуация, в которой оценивали качество гласных исследователи и аудиторы, была различной: исследователи слышали гласный в целом слове и могли находиться под влиянием морфонологических и орфографических ассоциаций; аудиторы же слышали изолированный гласный и гласный в соседстве с согласным, т. е. использовали только фонетическую информацию. Не исключено, что на исследователей могла влиять подсознательно и имеющаяся у них определенная гипотеза.

С помощью спектрального анализа можно было установить, какой же безударный гласный действительно произносится на месте орфографических *е, а, я*. Сравнение спектральных характеристик исследуемых гласных со спектральными характеристиками ударных |e|, |i| и безударного |i| показало, что фактически только 16% гласных из общего числа имеют характеристики, приближающие их к |e|, т. е. в действительности случаев «еканья» в исследуемом материале еще меньше.

Наиболее устойчивыми в области консонантизма оказались особенности произнесения отдельных слов, а именно произнесение слова *дождь* как |došt'|, *дождя* как |dažd'a| и произношение твердого согласного перед |e| в заимствованных словах. Устойчивыми эти особенности можно считать потому, что они встретились в произношении большинства испытуемых (более чем у 100 человек из 150). Остальные особенности в области консонантизма малоустойчивы, мера их сохранения — от 23 до 1%. Одни из них встречаются в небольшом числе случаев и у ограниченного числа дикторов (например, произношение |čt| в слове *что* характерно только для 13% испытуемых), другие особенности хоть и редко, но встречаются в произношении значительной части дикторов (например, произношение |t| в сложных сочетаниях согласных встречается более чем у 59% дикторов). Однако только 20% дикторов (28 человек) сохраняют эту особенность в 50 и более процентах; черта эта утрачивается.

Произношение |šč'| вместо |š':| встретилось лишь в 2% случаев. 87% испытуемых вообще не произносят |šč'| (следует заметить, что, судя по литературным данным, эта черта считалась обязательной в ленинградском произношении), у 13% эта черта встречается редко, и только один диктор в 50% случаев произносит |šč'| вместо |š':|.

Для проверки объективности полученных данных был проведен осциллографический и частично спектральный анализ материала. Сопоставление данных слухового и осциллографиче-

ского анализа показало, что в основном эти данные совпадают, однако имеются и некоторые расхождения: так, при осциллографическом анализе выявился ряд дополнительных случаев произношения |t| в сложных сочетаниях. Однако аудиторы, которым было предложено прослушать такие сочетания, |t| в них, как и при слуховом анализе, не услышали. Объясняется это, по-видимому, его небольшой длительностью (около 50—60 мс), небольшой смычкой (10—30 мс) и слабым взрывом (в тех случаях, когда |t| было отмечено аудиторами, его длительность была не менее 80 мс, длительность смычки — около 70 мс, взрыв — интенсивный).

Экспериментально-фонетическое исследование ленинградских особенностей произношения показало, что полной реализации «ленинградской нормы» мы не встретили ни у одного из дикторов, даже среди потомственных ленинградцев. Следует отметить, что даже те особенности, которые встретились в исследуемом материале, как правило, не проявляются последовательно: так, один и тот же человек может говорить |čto| и |što| |kan'ečna| и |kan'ešna|, |došt'| и |doš':|. Ряд особенностей уже не может считаться реально существующим в ленинградском произношении: почти никто не говорит |šč'i|, |šč'ot|, а в значительном большинстве случаев произносят |š':|.

В связи с этим важно исследовать, какой из вариантов кажется носителям языка (аудиторам) более правильным. Для решения этого вопроса тремя дикторами начитаны на магнитную ленту 11 типов вариантов: 1) |p'isoк| и [p'eⁱsoк] (всего 4 слова); 2) |žыlat'| и |žeᵘlat'| (4 слова); 3) |zvonк'ij| и |zvonкъj| (3 слова); 4) |vыtalк'ivat'| и |vыtalкъvat'| (3 слова); 5) |uč'ils'a| и |uč'ilsa| (4 слова); 6) |ats'uda| и |atsuda|; 7) |druк| и |drux| (3 слова); 8) |p'iš':a| и |p'išča| (13 слов); 9) |zv'er'| и |z'v'er'| (26 слов); 10) |kan'ešna| и |kan'ečna| (19 слов); 11) |l'ixкa| и |l'iккa| (9 слов).

Из записанного материала были составлены серии для опытов по восприятию. Аудиторам предъявлялись записи каждого диктора отдельно. Задание формулировалось следующим образом: выбрать тот из вариантов, который кажется правильнее, не объясняя причину выбора. В опытах участвовало 20 аудиторов, студентов II курса филологического факультета. Среди них были ленинградцы, москвичи и представители других городов бывшего Советского Союза. Первоначально данные их восприятия обрабатывались отдельно, но впоследствии, в связи с тем, что существенных различий в восприятии обнаружено не было, были объединены.

Анализ данных восприятия показал, что в большинстве случаев выбор одного из вариантов был достаточно определенным. Так, 85% аудиторов считают более правильным произношение мягких заднеязычных в прилагательных м. рода ед. числа |zvonк'ij|; 90% аудиторов выбирают вариант с мягким задне-

язычным в глаголах типа *выталкивать* |vыtalk'ivat'|; 97% аудиторов считают правильным произношение слова *отсюда* с мягким |s'| |atъs'uda|; 99% аудиторов выбирают вариант с |š':|; 96% аудиторов считают правильным произношение в конце слов *друг, снег, денег* глухого смычного |к|, а не щелевого |х|; 87% аудиторов выбирают вариант с первым твердым согласным в сочетаниях двух согласных со вторым мягким в тех случаях, когда эти согласные — разного места образования (слова типа *свить, зверь, две, лампе*).

В словах с сочетаниями согласных с |j| 69% аудиторов считают правильным вариант с первым твердым согласным (|vjot|). Эти результаты, по-видимому, свидетельствуют о расширении сферы действия общего процесса отвердения первого согласного в сочетаниях двух согласных со вторым мягким. Что же касается произнесения сочетания *чн* как |šn| или |č'n|, то выбор того или иного варианта связан с тем, о каком конкретном слове идет речь. Так, 86% аудиторов выбирают произношение |šn| в словах *конечно, двоечник, троечник, нарочно, яичница*; 92% аудиторов считают правильным произношение |c'n| в словах *молочный, булочная, порядочный, гречневый, яблочный, сливочный, огуречный, подсолнечный.*

Выбор вариантов произношения слов *пустячный, прачечная, перечница, подсвечник* оказывается менее определенным: 58% аудиторов считают правильным произношение |č'n| в этих словах, остальные |šn|.

Особого рассмотрения требуют данные выбора вариантов произношения |eˈ| или |i|, соответственно |eы| или |ы| на месте орфографических *е, а, я* после мягких и после твердых согласных. 54% аудиторов считают правильным произношение безударного |e| в позиции после мягких и 50% аудиторов — произношение безударного |e| в позиции после твердых согласных. Такие результаты, по-видимому, могут быть связаны с двумя факторами: обычным произношением дикторов и произношением самих аудиторов. Некоторая зависимость выбора аудиторов от обычного произношения диктора была обнаружена: для диктора № 1 характерно йкающее произношение, и аудиторы чаще выбирают у него как правильное произношение безударного |i| на месте орфографических *е, а, я*. Диктору № 3 свойственно éканье, и, по-видимому, в его произношении именно этот вариант звучит естественнее, красивее, чем не свойственное ему йканье, что и определяет выбор аудиторов. Зависимости данных восприятия таких случаев от произношения самих аудиторов обнаружить пока не удалось, вероятно, необходимо провести более тщательный анализ произношения аудиторов и расширить материал анализа. Однако подобная методика анализа произношения представляется перспективной, хотя интерпретация полученных в результате опытов по восприятию данных оказывается довольно сложной и требует под-

робного исследования орфоэпических и орфофонических особенностей произношения и дикторов, и аудиторов.

Подробный экспериментально-фонетический анализ произношения исследованной группы дикторов показал уже отмечающуюся в ряде работ общую тенденцию к выравниванию территориальных вариантов русского литературного языка, к сглаживанию различий между ними, к формированию некоей общей произносительной нормы, объединяющей черты московского и ленинградского вариантов.

Л. В. Щерба в работе «О нормах образцового русского произношения» писал: «...в произношении будущего будет отменено все чересчур местное, московское или ленинградское, орловское или новгородское, не говоря уже о разных отличительных чертах других языков, вроде кавказского или среднеазиатского „гортанного" *х*, украинского *г*, татарского *ы* и т. д.» (Щерба, 1957, с. 111). Вот именно этот процесс образования единой произносительной нормы, лишенной местных черт, происходит в течение последних 30—40 лет. Какое место в этом выравнивании территориальных вариантов русского литературного языка занимает сознательная нормирующая деятельность, а какое — живые миграционные процессы, предстоит еще выяснить при исследовании различных аспектов развития литературного языка.

Образование единой произносительной нормы, заимствовавшей частично особенности московского варианта (йканье (ыканье), произнесение |š':|, произнесение |št|, |šn| вместо |č't|, |č'n| в ряде слов и т. п.), частично ленинградского варианта (произнесение мягкого |s'| в возвратных частицах глаголов, мягких согласных перед флексией в прилагательных м. рода ед. числа |t'ix'ij| вместо |t'ixъj| и т. п.) определяется фонологической системой русского языка. Изменения произношения происходят в тех случаях, когда не затрагивается фонологическая система как таковая. Так, в русском языке фонемы |к| и |к'| противопоставлены, системе одинаково присущи сочетания |к'i| и |къ| |гак'i| и |гакъм'i|; именно поэтому и мог произойти сдвиг в соотношении произносительных вариантов. Ленинградский вариант, требующий произношения |к'|, становится основным, ведущим. Не противоречит системе и произношение |s'| на конце слова |bros'|, поэтому возможно распространение произношения в возвратных частицах (|nad'ejus'|). Чередование твердого и мягкого согласных в сочетаниях двух согласных со вторым мягким не затрудняет понимание слова; распространение произношения твердого согласного перед последующим мягким, таким образом, не противоречит системе языка. Перспективными могут оказаться, в первую очередь, те изменения, которые не вступают в конфликт с фонологической системой.

Исследование записей речи 220 дикторов, жителей Москвы

и Ленинграда, различающихся возрастом (от 20 до 60), образованием (среднее и высшее) и временем проживания в указанных городах (от 5 лет и более), позволяет судить об основных чертах современной русской орфоэпии и орфофонии.

Среди дикторов были и потомственные москвичи и ленинградцы. 20 дикторов из 220 были записаны после предварительного слухового анализа их речи: были отобраны те носители языка, в речи которых не было обнаружено никаких диалектных, просторечных особенностей, указывающих на принадлежность к узкоограниченной социальной группе. Эти 20 дикторов составили контрольную группу; результаты анализа их произношения постоянно сравнивались с результатами анализа остальных дикторов (200 человек).

Основной орфоэпической особенностью вокализма является йканье (ыканье), т. е. произношение в безударной позиции на месте орфографических *е, а, я* — гласного |i|. Следует заметить, что еканье еще встречается (в 15,5% у ленинградцев, в 10,3% у москвичей), но оно постепенно вытесняется йканьем. В контрольной группе дикторов еканье вообще не было отмечено.

В начале слов на месте орфографических *е, я* произносится |ji| или |i|, т. е. *Европа*, *Япония* как |jivropa| или |ivropa|, |jipon'ija| или |ipon'ija|. На месте орфографического *э* в начале слов (исследовалось 32 слова) произносится |ы| или |е|. Опыты по восприятию показали, что исследуемый гласный в 55% случаев определялся аудиторами как |ы|, в 45% — как |е| (при этом часто указывалось, что |е| похож на |ы|). Произношение |ы| или |е| на месте *э* оказалось связанным с частотой употребления слов в речи (чем выше частота, тем вероятнее произнесение |ы|) и с позицией гласной по отношению к ударению (чем дальше от ударения, тем чаще |ы|). Сопоставление чтения и спонтанной речи показало, что |ы| в указанных случаях в спонтанной речи встречается чаще. Произношение начального |ы| является, по-видимому, результатом широко распространяющихся йканья (ыканья), тенденцией вообще не произносить в безударных слогах |е|. В то же время произношение безударного |ы| может, как нам представляется, обеспечиваться и системными факторами: наличием самостоятельной фонемы |ы|; именно система дает норме такую возможность произнесения.

Исследование безударных гласных в заударных флексиях показало, что общий характер редукции заударных флексий такой же, как и остальных заударных гласных. Во флексии, так же как и в основе, наибольшей редукции подвергается гласный |а|. Спектральный анализ гласных в заударных флексиях показал, что и на уровне акустических характеристик нельзя говорить о большей сохранности заударных флексий по сравнению с остальными заударными гласными.

Исследование произношения существительных ж., м., и ср. рода ед. числа дат. и пр. падежей показало, что говорить о сохранении безударной |е| в заударных флексиях нет оснований (*в поле* произнесено |ї pol'i|, *в Польше* |ї pol'šы| и т. п.).

В глаголах ед. и мн. числа 3-го лица с твердой основой бо́льшая часть дикторов произносит редуцированное |ы| (*кружит* |kružыt| и *кружат* |kružыt|), с мягкой основой — редуцированное |i| (*медлит* |m'edl'it| и *медлят* |m'edl'it|).

Фонетическая оценка заударных флексий имен прилагательных оказывается более сложной. Специальные опыты выяснения роли смыслового контекста при оценке форм слов (прилагательное *добрая* склеивалось с существительным *утро*, прилагательное *доброе* — с существительным *слава*; аудиторы не замечали неестественности звучания и опознавали род по существительному) показали преобладающую роль контекста в случае фонетической неопределенности заударного гласного. Можно предполагать, что в естественных условиях форма слова определяется не столько по ее фонетическим характеристикам, сколько на основе анализа смысловых отношений.

Формы прилагательные ед. числа ж. и ср. рода с твердой основой фонетически реализуются как |dobrai|, с мягкой — как |s'in'ii|; мн. числа с твердой основой — как |dobrыi|, а с мягкой — как |s'in'ii|.

Особого рассмотрения требует произнесение безударной |а| после шипящих, так как согласно старомосковской норме на месте *а* произносилось |ы|. Слуховой и спектральный анализ показал, что в большинстве слов (например, *жара, шалун, шампанское, шаги* и т. п.), за исключением слов *лошадей, жалеть* (и всех форм этого слова), *жакет, ржаной,* произносится |а| соответствующей степени редукции.

Исследование спектральных характеристик безударных гласных показало, что в небольшом числе иностранных по происхождению слов сохраняется безударное |о| (в среднем в 12,5%). Вероятность произнесения безударного |о| (было исследовано произношение 76 иностранных по происхождению слов) зависит, во-первых, от частоты употребления слова в речи (в словах типа *шоссе, бордо, тоннель, поэт* только в 5% случаев произносится |о|, в словах типа *вето, досье* — в 30%), во-вторых, от позиции безударного |о| в слове: чаще |о| сохраняется в тех случаях, когда стоит рядом с гласным (*оазис, боа* — 40%), особенно если безударный |о| находится в абсолютном конце слова (в словах *какао, радио, адажио* в 90% случаев было произнесено |о|).

При исследовании орфоэпических особенностей в области консонантизма особое внимание было уделено произношению мягких согласных. В ряде случаев система языка дает норме две возможности произнесения: твердый и мягкий согласный. Так, в прилагательных м. рода ед. числа и в возвратных частицах гла-

голов в течение долгого времени одинаково возможным было произнесение и твердых, и мягких согласных: |t'ix'ij| и |t'ixъj|, |učus'| и |učus|, так как система допускает |k'i| — |v'ilk'i| и |kъ| — [v'ilkъj], |s'| — |trus'| и |s| — |trus|. Все исследованные нами дикторы во всех случаях в прилагательных м. рода ед. числа произнесли мягкий заднеязычный, в возвратных частицах глаголов — в 80% случаев.

В иностранных по происхождению словах тоже возможно употребление и мягких, и твердых согласных перед гласным переднего ряда |e|. Теоретически были основания ожидать в иностранных словах, широко бытующих в русском языке, произнесения мягких согласных перед |e|: |t'emp| и даже |t'ez'is|. Однако этого не произошло именно потому, что в системе была заложена возможность произнесения твердого согласного перед |e| в русских словах: |šest|, |žest|, |šes't'|. Согласно современной произносительной норме перед гласным переднего ряда |e| в словах указанного типа могут произноситься и мягкие, и твердые согласные. Экспериментально-фонетическое исследование 103 слов показало, что имеется определеная связь между качеством согласного и употреблением мягкого или твердого согласного. Так, в 80% слов заднеязычные — мягкие (*трахея* |trax'eja|, *схема* |sx'ema|), в 75% губные — мягкие (*дефект* |dif'ekt|, пропеллер |prap'el'ir|); переднеязычные же (за исключением |l|, которое всегда мягкое) в 75% твердые (*темп* |temp|, *тезис* |tez'is|, *фонетика* |fanet'ika|), по-видимому, в первую очередь, в связи с тем, что в системе заложена возможность сочетания с гласным |e| именно переднеязычных.

Рассматривая сочетания согласных, следует отметить общую тенденцию к распространению сочетания твердый+мягкий согласный вместо мягкий+мягкий; первый мягкий согласный сохраняется лишь в сочетаниях двух переднеязычных[3], например |s't'|, |z'd'|, |n't'|, |n'c'|, |n's'|: |s't'ep'|, |z'd'es|, |v'in't'ik|, |v'enč'ik|, |p'en's'ija|.

В сочетаниях двух согласных, из которых второй — |j|, первый согласный преимущественно твердый (*объем* |abjom|, *разъезд* |razjest|); только в сочетаниях |t'j, d'j, r'j, l'j, n'j| (на стыках корня и суффикса) первый согласный — мягкий (|stat'ja, sud'ja, kamar'jo, kol'ja, pavl'in'ja|).

Возможность употребления то твердых, то мягких согласных перед последующим мягким и связь этого процесса с фонетическими характеристиками звуков сочетания объясняется отсутствием в системе языка ограничений сочетания разных по твердости — мягкости согласных.

---

[3] Было проведено специальное исследование всех возможных в русском языке сочетаний двух согласных со вторым мягким; спектральный анализ и анализ восприятия носителей языка показал, что в 216 сочетаниях из 298 дикторами был произнесен твердый согласный перед последующим мягким.

Исследование произношения сочетания *чт* показало, что в слове *что* и его производных произносится в 100% случаев |št|. Что же касается произношения сочетания *чн*, то в ряде слов (например, *конечно, скучно, яичница, пустячный, двоечник, троечник*) произносится в большинстве случаев |šn|, в ряде слов (например, *булочная, молочная, яблочный, порядочный*) — |č'n|. Анализ произношения слов с сочетанием *чн* показал определенную роль морфонологического фактора в выделении ведущего варианта нормы: из всех слов, в которых *чн* произносится как |č'n|, в 77% случаев в корне — *к* (т. е. здесь, по-видимому, чередование *к*‖*ч*), в 9% — *ч*, в 14% — *ц* (*булка — булочная* |bulač'naja|, *порядок — порядочный* |par'adač'nыj| и т. п.).

Широко известно, что в конце слов перед паузой в русском языке не могут произноситься звонкие согласные: они в этих случаях чередуются с глухими. Это относится и к согласному *в*, несмотря на сложные фонетические характеристики этого согласного, т. е. слова *лов, без пределов* и т. п. должны произноситься как |lof|, |b'es pr'id'elaf|.

На стыках знаменательных слов (при отсутствии пауз) звонкий согласный в конце первого слова произносится только перед последующим звонким шумным (*город большой* |gorad bal'šoj|), в остальных случаях произносится глухой согласный (*город солнечный* |gorat soln'ičnыj|, *город летний* |gorat l'etn'ij|, *город великий* |gorat v'il'ik'ij|, *город огромный* |gorat agromnыj|).

На стыках служебного и знаменательного слов звонкий согласный произносится на конце служебного слова не только перед последующим звонким шумным, но и перед сонантом и гласным (*через год* |čir'iz got|, *через миг* |č'ir'iz m'ik|, *через век* |č'ir'iz v'ek|, *через огонь* |c'ir'iz agon'|).

При рассмотрении орфофонических характеристик гласных[4] следует, в первую очередь, обратить внимание на произнесение гласного |a|.

Гласный |a| в 1-м предударном слоге не только более краткий, чем в ударной позиции, но и более закрытый, а в ряде случаев и более задний; во 2-м предударном слоге — более закрытый, чем в 1-м. В заударном неконечном слоге гласный еще более краткий, а иногда и более задний, чем во 2-м предударном. Степень закрытости гласного и его отодвинутости назад, а также его сокращения в безударной позиции может быть разной: так, сравнение длительности безударных гласных выявило большее сокращение гласного |a| в позиции 2-й степени редукции у дикторов-москвичей по сравнению с дикторами-ленинградца-

[4] Безударные /ы/, /i/, /u/ по сравнению с соответствующими ударными изменяются, главным образом, количественно (Бондарко и др., 1966).

ми; сравнение частотных значений формант показало, что эти же гласные в московском произношении и более закрытые, и более задние, чем у ленинградцев.

Орфографическим *щ, сч, зч* соответствует в произношении всех дикторов |š':|, а не |šč|, т. е. |š':i|, а не |šč|. Говоря о других орфофонических особенностях согласных, целесообразно отметить те, нарушение которых допускается фонологической системой русского языка. Речь идет о согласных |š|, |ž| и |c|, которые должны произноситься как твердые, хотя в системе и отсутствует противопоставление твердых |š|, |ž|, |c| соответствующим мягким; |č|, который должен произноситься как мягкий. Следует еще обратить внимание на необходимость произнесения согласного *г* как смычного, а не фрикативного, несмотря на отсутствие в русском языке противопоставления |g|—|γ|.

Таковы лишь основные черты современной орфоэпии и орфофонии. Реальные пути развития произносительной нормы определяются фонологической системой языка и прежде всего существующей в ней вариантностью. Произносительная норма также характеризуется известной вариантностью, которая разрешена системой.

При утверждении того или иного орфоэпического или орфофонического варианта в качестве ведущего варианта нормы большое значение имеет анализ основных тенденций развития фонологической системы. Чаще побеждают варианты, соответствующие тенденциям развития системы.

Знание основных направлений развития фонологической системы имеет особое значение для кодификаторов, вынужденных сознательно выбирать один из вариантов, в тех случаях, когда один становится ведущим вариантом нормы, а второй — отживающим. При выборе одного из вариантов кодификатором необходимо учитывать разнообразные фонетические, фонологические, морфонологические и экстралингвистические факторы, опираться на экспериментально-фонетические исследования реальной речи носителей языка.

Литература

Аванесов Р. И. 1) О нормах русского литературного произношения // Рус. язык в школе. 1961. № 6; 2) Русское литературное произношение. М., 1972.

Баринова Г. А., Ильина Е. Е., Кузьмина С. М. О том, как проверяли вопросник по произношению // Развитие фонетики современного русского литературного языка: Фонолог. подсистемы. М., 1971.

Бельчиков Ю. А. О нормах литературной речи // Вопросы культуры речи. Вып. 6. М., 1970.

Бондарко Л. В., Вербицкая Л. А., Зиндер Л. Р. Акустические характеристики безударности // Структурная типология языков. М., 1966.

Бондарко Л. В., Вербицкая Л. А. О фонетических характеристиках заударных флексий в современном русском языке // Вопр. языкозн. 1973. № 1.

Булаховский Л. А. Курс русского литературного языка. 5-е изд. Т. 1. Киев, 1952.

Гельгардт Р. Р. К вопросу об орфоэпической норме в русском литературном языке // Рус. язык в школе. 1946. № 5—6.

Голанов И. Г. Основные положения русской орфоэпии // Учен. зап. Моск. пед. ин-та. 1960. № 148, вып. 10.

Грот Я. И. 1) О некоторых особенностях в системе звуков русского языка // ЖМНПР. 1852. Т. 74; 2) Филологические разыскания. Т. 2. СПб, 1869.

Ицкович В. А. 1) Языковая норма. М., 1968; 2) Норма и ее кодификация // Актуальные проблемы культуры речи. М., 1970.

Косериу Э. Синхрония, диахрония и история // Новое в лингвистике. Вып. 3. М., 1963.

Матезиус В. Основные принципы культуры языка // Пражский лингвистический кружок. М., 1967.

Петрищева Е. Ф. К вопросу о критериях нормативности // Вопросы культуры речи. Вып. 8 М., 1967.

Русский язык по данным массового обследования: Опыт социолингв. изучения. М., 1974.

Семенюк Н. Н. Проблема формирования норм немецкого литературного языка: XVIII столетие. М., 1967.

Скворцов Л. И. Норма литературного языка: Культура речи // Актуальные проблемы культуры речи. М., 1970.

Соссюр Ф. де. Курс общей лингвистики. М., 1933.

Утехина Н. П. О некоторых орфоэпических нормах в современном русском языке // Рус. язык в школе. 1964. № 6.

Ушаков Д. Н. Русская орфоэпия и ее задачи // Русская речь. Вып. 3. М., 1928.

Чернышев В. И. 1) О произношении // Рус. филол. вестник (Варшава). 1911. № 2; 2) Законы и правила русского произношения. Пг., 1915. С. 12—13. — То же // Избр. труды. Т. 1. М., 1970; 3) Псковское наречие // Труды комиссии по рус. языку. 1931. Т. 1; 4) Как говорят в Петербурге // Избр. труды. Т. 2. М., 1970.

Щерба Л. В. 1) Русские гласные в качественном и количественном отношении. СПб., 1912; 2) Спорные вопросы русской грамматики // Рус. язык в школе. 1939. № 1; 3) К вопросу о русской орфоэпии // Избр. работы по рус. языку. М., 1957; 4) О нормах образцового русского произношения // Там же; 5) О трояком аспекте языковых явлений и об эксперименте в языкознании // Языковая система и речевая деятельность. М., 1974.

Schindler F. Beiträge zur deutschen Hochlautung. Hamburg, 1974.

МЕТОДИКА ОБУЧЕНИЯ РУССКОМУ ЯЗЫКУ КАК РОДНОМУ В СРЕДНЕЙ ШКОЛЕ

Методика обучения тому или иному предмету должна обязательно содержать ответы на три главных вопроса: 1) о содержании обучения (чему учить?), 2) о целях обучения (для чего учить?) и 3) о принципах, методах и приемах обучения (как учить?). Представляется, что для характеристики методики обучения родному языку в школе как отрасли прикладного языкознания необходимо прокомментировать в первую очередь ее положения о содержании и целях обучения. Эти положения методики обусловлены прежде всего спецификой самого учебного предмета «русский язык» по сравнению с другими дисциплинами в средней школе, а также связями методики обучения родному языку со смежными науками — лингвистикой, психологией и педагогикой. О специфике русского языка как учебного предмета говорится в программе восьмилетней и средней школы, в учебнике по методике и в пособиях для учителей. «Родной язык является не только предметом изучения в школе (т. е. его не только изучают), но и средством обучения (т. е. с его помощью учащиеся изучают все другие школьные предметы). В этом его специфика» (Текучев, 1980, с. 8; см. также: Методика преподавания..., 1990, с. 22). Дело в том, что родной язык является важнейшим средством общения в национальном коллективе, в частности, важнейшим средством передачи и получения самой разной информации.

Мышление ребенка развивается одновременно с процессом овладения им родным языком, усвоения форм родной речи. Известное положение о диалектической связи языка и мышления обязывает учителя-словесника постоянно заботиться о развитии логического мышления учащихся. Предметом изучения на уроках русского языка в школе являются различные значимые единицы языка — предложения, словосочетания, слова и морфемы, в которых зафиксированы, материализованы резуль-

таты познавательной деятельности людей. В связи с этим различным упражнениям — и лексическим, и грамматическим — может быть и должна быть придана форма собственно логических упражнений. В результате многократных упражнений обучающийся все более четко и более глубоко начинает выражать свои мысли в понятиях, суждениях и умозаключениях.

Постепенно овладевая на уроках русского языка литературными нормами, обогащая свою речь, совершенствуя ее точность, выразительность и таким образом развивая свой интеллект, учащиеся приобретают «возможность успешно изучать все другие учебные предметы и тем самым расширять свои знания об окружающей действительности» (Программы..., 1985, с. 1; см. также: Программы для общеобразовательных учреждений, 1992, с. 46). Эта мысль еще более конкретно выражена в пособии для учителей Л. П. Федоренко: «Пока учащиеся не овладеют в достаточной мере родным языком, они не усвоят никаких знаний на уроках физики, естествознания и т. д., воспитательные возможности уроков истории, литературы останутся для них нереализованными» (Федоренко, 1973, с. 14).

Вопрос о специфике предмета «родной язык», о его месте в учебном плане школы рассматривается также в работах специалистов по преподаванию других школьных дисциплин. Особенности предметов «родной язык» и «иностранный язык», их место в ряду других школьных дисциплин находятся в центре внимания авторов учебных пособий по обучению иностранным языкам в школе (Методика..., 1982, с. 37—40; Теоретические основы..., 1981, с. 15—17; Общая методика..., 1967, с. 5—18). В этих пособиях говорится, что разные предметы школьного учебного плана неоднородны с точки зрения задач обучения им. Большую группу предметов составляют естественнонаучные и гуманитарные дисциплины (химия, физика, биология, история и др.). На уроках по этим предметам изучаются основы соответствующих наук. Главной задачей обучения этим предметам является сообщение учащимся научных знаний, формирование у них научных понятий и представлений. Практический аспект обучения, формирование умений и навыков играет при этом явно подчиненную роль по сравнению со знаниями. Вторую группу учебных дисциплин составляют предметы эстетического цикла (музыка, изобразительное искусство, литература). Основная функция предметов этой группы — воспитательная. Главной задачей предметов третьей группы (трудовое обучение, физическое воспитание) является формирование конкретных умений и навыков, связанных с физической деятельностью человека.

Особую, четвертую, группу составляют предметы «русский (родной) язык» и «иностранный язык». Принципиальное отличие их от дисциплин первой и второй групп заключается в том, что усвоение знаний о языке не является самоцелью, а пред-

ставляет собою средство для формирования навыков и умений, необходимых для оптимального использования языка как средства общения. Главной, преобладающей в обучении языкам является коммуникативная функция. От третьей группы дисциплин родной и иностранный языки отличаются тем, что формируемые речевые умения — это интеллектуальные умения, а не навыки физической деятельности.

Вместе с тем, как пишут авторы упомянутых пособий, несмотря на значительную общность учебных предметов «родной язык» и «иностранный язык» между ними существуют и различия, которые сводятся к разному соотношению коммуникативной и образовательно-воспитательной функций: «При обучении родному языку коммуникативная и образовательно-воспитательная функции по сути дела равноценны или почти равноценны» (Методика.... 1982, с. 38); «Отличие учебного предмета „иностранный язык" от учебного предмета „родной язык" заключается в том, что основной и ведущей целью в преподавании иностранных языков является коммуникативная цель, в то время как в преподавании родного языка эта цель не является главной» (Общая методика..., 1967, с. 13).

С точки зрения современных требований, предъявляемых обществом к изучению родного языка в школе, с таким утверждением нельзя согласиться. Коммуникативность в обучении в очень большой степени свойственна обоим предметам. Различия в содержании и целях обучения родному и иностранному языкам связаны с кругом и характером формируемых речевых навыков и умений.

На уроках русского (родного) языка должны быть сформированы навыки и умения, более разнообразные и более высокой степени сложности, чем при обучении иностранному языку. В этом отношении отраженное в действующей программе содержание обучения родному языку необходимо уточнить, обогатить путем значительного расширения круга литературных норм, подлежащих усвоению.

Речь выпускников средней школы должна быть правильной, богатой и выразительной, а между тем очень многие воспитанники школы не владеют в достаточной степени нормами произношения и ударения, словоупотребления, образования грамматических форм слов, построения словосочетаний и предложений (из-за бедности словарного и грамматического запаса речи).

Важнейший вопрос о содержании занятий по русскому языку, родному языку учащихся, может быть решен на основании связи методики с лингвистической наукой. Содержание обучения родному языку во многом отличается от содержания обучения иностранному языку, так как дети еще до школы в силу потребностей жизни, благодаря языковому окружению овладевают путем подражания нормами своего родного языка — орфоэпическими, лексическими, грамматическими. В связи с этим

может даже возникнуть сомнение в необходимости обучать русскому языку русских учащихся.

Однако в какой степени ребенок владеет родным языком и, в частности, высшей его формой — литературным языком до изучения его в школе? Совершенно очевидно, что без специального обучения в школе ребенок не может достаточно полно овладеть литературным языком, разнообразными его нормами.

Вопрос о содержании обучения русскому языку как родному может быть решен прежде всего в свете лингвистического учения о литературном языке, о его сложной системе функциональных стилей, о диалогической и монологической формах речи, на основе положений трудов проф. А. М. Пешковского и акад. Л. В. Щербы. Нормы книжной речи с ее максимально обогащенным словарем и максимально усложненным синтаксисом усваиваются только в процессе специально организованного школьного обучения, важнейшей частью которого является изучение грамматики. Книжная разновидность литературного языка должна быть предметом школьного изучения, так как ею нам приходится постоянно пользоваться в условиях сложной современной культурной жизни, как только появляется необходимость говорить или писать на отвлеченные темы.

Диалогической формой речи, умением обмениваться репликами с собеседником человек овладевает, в основном, и без школьного обучения. Иначе обстоит дело с нормами монологической формы речи, связного развернутого высказывания, рассказа, адресованного многим слушателям (читателям). Нормы, правила построения монолога должны постоянно изучаться на уроках русского языка.

Можно говорить о трех основных разделах содержания обучения русскому языку как родному. I. Развитие речи учащихся — богатое по содержанию понятие методики, предполагающее работу учителя в трех направлениях: 1) совершенствование навыков владения нормами литературного языка в его нейтральном и разговорном стилях, постоянная забота учителя о правильности речи учащихся, о преодолении просторечного (иногда — диалектного) произношения, словоупотребления и построения предложений; 2) обогащение речи учащихся как новыми, не знакомыми для них словами и фразеологическими оборотами, так и изучаемыми синтаксическими конструкциями (речь детей бедна и в лексическом, и в грамматическом отношении); 3) формирование навыков связной речи, усвоение норм монологической формы речи как конечная цель изучения родного языка в школе. II. Формирование навыков правописания — орфографических и пунктуационных. III. Изучение основ науки о русском языке, имеющее, главным образом, прикладное, практическое значение.

Главной целью обучения в школе русскому языку как родному является практическая, коммуникативная. Определение

содержания и целей обучения родному языку предполагает ярко выраженную практическую направленность изучения русского языка. Конечной целью обучения родному языку является выработка различных практических навыков и умений. Школа должна научить «правильно читать, писать и говорить, а также мыслить» (Д. Н. Ушаков).

Что касается принципов обучения и выбора наиболее целесообразных, эффективных и экономных методов и приемов передачи знаний и формирования умений и навыков, то в значительной степени их определение и описание обусловлены связями методики с психологией (особенно с детской психологией, психологией подростка) и педагогикой (главным образом, с дидактикой — теорией обучения).

Литература

Методика обучения иностранным языкам в школе. М., 1982.

Методика преподавания русского языка / Под ред. М. Т. Баранова. М., 1990.

Общая методика обучения иностранным языкам в школе. М., 1967.

Программы для общеобразовательных учреждений: Русский язык. М., 1992.

Программы восьмилетней и средней школы: Русский язык. М., 1985.

Текучев А. В. Методика русского языка в средней школе. 3-е изд., перераб. М., 1980.

Теоретические основы обучения иностранным языкам в школе. М., 1981.

Федоренко Л. П. Принципы обучения русскому языку. М., 1973.

АВТОМАТИЗИРОВАННЫЕ ОБУЧАЮЩИЕ СИСТЕМЫ

Компьютеризация обучения представляет собой одну из чрезвычайно сильных тенденций в современной педагогике. Вместе с тем это одна из сложнейших задач, стоящих перед системой образования. Именно автоматизированное обучение способно коренным образом повысить эффективность обучающих воздействий благодаря повышенной мотивации, одновременно сокращая затраты человеческого труда вследствие того, что рутинные операции, связанные с образовательным процессом, перекладываются на «плечи» вычислительной техники и, следовательно, выполняются человеком лишь один раз, повторяясь машиной столько раз, сколько необходимо.

Особую актуальность автоматизация приобретает в связи с появлением персональных компьютеров, отличающихся относительной дешевизной при возможности исключительно эффективной работы в диалоговом режиме.

Разумеется, автоматизация обучения отнюдь не означает, что труд педагога становится ненужным. Однако автоматизация позволяет избавить последнего от такой деятельности, которая не требует от человека никакого творчества, сводясь к набору элементарных механических операций, и с которой ЭВМ может справиться с большей эффективностью, ибо, в отличие от человека, не подвержена усталости и может выполнять их существенно быстрее.

Процесс автоматизации обучения состоит в разработке и применении автоматизированных обучающих систем (АОС). Часто в понятие АОС включают комплекс технических средств обучения и пакет обучающих программ (или даже одну такую программу), каждая из которых рассчитана на одно занятие, т. е. на один сеанс взаимодействия с обучаемым. Такой подход является односторонним: само понятие обучающей системы подсказывает, что конкретные обучающие программы должны

составлять целостный комплекс не только с точки зрения их соотнесения с учебным планом, но и с точки зрения внутренней организации АОС. Такое объединение может иметь два основания: общность дидактического материала, рассматриваемого как данные, используемые программами (базы данных по предметной области, или базы знаний в широком смысле слова), и использование сведений о предыдущей деятельности обучаемых при работе в рамках данной АОС (базы данных по результатам учебной деятельности) наряду с наличием в системе средств настройки на конкретного обучаемого с учетом результатов его предшествующих взаимодействий с данной обучающей системой, т. е. автоматического выбора стратегии и тактики обучения.

Не следует, впрочем, полностью отрицать тот факт, что и одиночная обучающая программа способна приносить определенную пользу, однако надо отдавать себе отчет в том, что эффективность той же самой программы, входящей в состав АОС, становится на порядок более высокой. Поэтому разрабатываться такие программы должны с учетом их возможного будущего взаимодействия друг с другом в рамках единой системы.

Таким образом, АОС может в общем случае включать в себя следующие основные компоненты: 1) комплекс технических средств обучения; 2) пакет обучающих программ, 3) базу знаний о предметной области; 4) базу данных, регистрирующую результаты обучающих воздействий.

АОС могут использоваться при обучении различным дисциплинам. Однако здесь мы будем говорить в основном о системах, предназначенных для обучения иностранным языкам.

Функции АОС определяются теми обучающими воздействиями, которые программы, входящие в систему, оказывают на обучаемых. К числу основных функций этих программ относятся:

— собственно обучение;
— тренировка;
— контроль;
— оценка.

Такое разделение чрезвычайно существенно при проектировании и разработке обучающих программ: оно определяет общий вид сценария программы, форму диалога с обучаемым (если диалог предусмотрен), а также в некоторой степени и действия обучающего как до, так и после сеанса взаимодействия обучаемого с программой.

Поскольку одним из основных принципов автоматизированного обучения можно считать его целостность, т. е. способность к реализации законченного фрагмента обучения, в АОС необходимо наличие обучающих программ всех трех видов (информирующие, тренировочные, контрольные), ибо лишь в этом

случае возможна такая реализация. Однако следует сказать, что целесообразность попыток совместить в рамках одной обучающей программы разные функции вызывает серьезные сомнения в силу принципиальных различий в подзадачах, решаемых алгоритмами, предназначенными для выполнения различных функций, и совмещение в рамках одной программы этих алгоритмов ведет к эклектичному результату.

Что касается дидактического материала, то он может быть одинаковым для всех трех видов программ; они различаются лишь способом его использования, что является важным аргументом в пользу независимости материала (структуры данных, используемых программой) от самой программы. Одни и те же данные, находящиеся, например, в базе знаний или записанные на универсальном языке представления (под которым понимается совокупность средств, позволяющих формализованно представить данные и методы их обработки), могут по-разному использоваться программами, выполняющими разные функции. Помимо очевидной экономии человеческого труда это дает возможность для однократного внесения аналогичных и непротиворечивых изменений в материал, используемый разными программами.

Информирующие программы осуществляют предъявление дидактического материала обучаемым, или информирование, что требует особого оформления данных (раскадровки, размещения информации на экране, выделения особо важной информации). Желательно использование графики, музыкальных средств компьютера и других приемов, позволяющих оживить достаточно монотонный процесс восприятия информации в текстовом режиме, т. е. в конечном счете повысить мотивацию. Необходимо предоставить обучаемому возможность при необходимости вернуться к уже предъявленному фрагменту информации.

Тренировочные программы требуют несколько иного подхода к своей структуре: желательно обеспечить обращение обучаемого к базе знаний, если таковая имеется, или использовать некоторую систему подсказок, затрагивающих наиболее трудные части материала и наиболее типичные ошибки: кроме того, не лишним будет диалог обучаемого с программой и обратная связь. При выполнении такой программы желательно обеспечить также информирование обучаемого в том случае, когда он допускает ошибки (в простейшем случае — информация о факте допущенной ошибки, в более сложных программах — попытка показать, в чем состоит эта ошибка).

При работе с *контролирующими* программами имеет смысл хранить результаты выполнения каждого задания в базе данных по результатам учебной деятельности. Кроме того, особо тщательно следует различать реальные и случайные ошибки, что имеет весьма большое значение в лингвистических обучаю-

щих программах, в которых особый вес приобретает вопрос об опечатках.

В случае, когда в программе реализуется функция *оценки*, необходим дальнейший анализ ошибок, определение степени их «тяжести» и статистическая обработка результатов выполнения заданий.

Возможно выделение нескольких дополнительных функций, выполняемых АОС. Эти функции реализуются факультативно, в дополнение к одной или нескольким из основных. К числу дополнительных функций АОС можно отнести следующие:

— *регистрацию результатов учебной деятельности*. Особое значение эта функция приобретает в АОС, имеющих в своем составе соответствующую базу данных; более того, при наличии такой базы данных регистрация результатов обучающих воздействий является наиболее эффективным средством ее ведения;

— информирование обучаемого. Эта функция может быть реализована в полном объеме при наличии в системе базы знаний: в этом случае обучаемому может быть предоставлена возможность посредством обращения к ней через запросы получать необходимую ему информацию;

— ведение диалога. Для ведения диалога необходим хорошо продуманный, приближающийся к естественному язык, позволяющий строить и распознавать необходимое множество высказываний;

— выбор стратегии и тактики обучения (настройка системы *на обучаемого*). *Для такой настройки* целесообразно использовать базу данных по результатам учебной деятельности.

Распределение этих дополнительных функций по различным видам обучающих программ различно. Так, реализация функции регистрации результатов возможна только в контролирующих программах, а функция информирования в таких программах представляется совершенно излишней, если не сказать вредной.

Рассмотрим ряд прикладных проблем, связанных с построением обучающих систем, и сформулируем ряд принципов, соблюдение которых может оказаться полезным в этой деятельности.

1) Недостаточно высокая эффективность многих АОС порождается отсутствием *комплексного подхода* к обучению. Под комплексностью обучения мы понимаем то, что применение ЭВМ должно пронизывать весь его процесс. Оно может не быть эффективным, если обучающие программы носят эпизодический характер, а учебный материал, охватываемый ими, включает в себя лишь немногие разделы курса обучения. Скорее наоборот, не исключено, что временное повышение мотивации и другие положительные стороны автоматизации обучения спо-

собны вызвать обратный эффект в тех разделах курса, которые автоматизированы не были, «плюсы» превратятся в «минусы».

2) Обучение языку должно быть *многоаспектным*. Это означает, что целесообразно такое построение обучающей программы, когда информация, составляющая содержание обучения, относится к разным уровням языка. При этом сам процесс обучения становится менее монотонным; кроме того, у обучаемого формируется представление не только о системе языка в ее статическом аспекте, но и о функционировании различных единиц, если, к примеру, усваивая новые номинативные единицы, он имеет возможность сразу использовать их в различных синтаксических конструкциях. Необходимо также заметить, что, только получая сведения о синтаксическом использовании номинативных единиц, обучаемый может адекватно усвоить такие их свойства, как валентность и сочетаемость.

3) Принцип *естественности* обучения включает в себя, во-первых, тщательный отбор учебного материала, демонстрирующего наиболее представительные и частотные явления изучаемого языка; во-вторых, в программах, допускающих управление со стороны обучаемого (например, обращение к базе знаний), необходима такая организация диалога с машиной, которая не требует никаких специальных навыков работы с компьютером; в-третьих, и сам диалог с программой должен вестись на максимально приближенном к естественному языке.

Последнее требует некоторого уточнения. Дело в том, что, как правило, во многих системах (не только обучающих) для целей организации диалога используются англоподобные языки, что не вызывает особых затруднений у большинства людей, имеющих соответствующую подготовку (а пока с компьютерами имеют дело, в основном, именно такие люди). Однако в обучающих языковых программах, особенно в тех, которые рассчитаны на начинающих, имеет смысл использовать только языки, построенные на основе родного языка обучаемого, не говоря уже о том, что самым распространенным в качестве изучаемого как раз и является английский, и использование его в качестве интерфейса предполагает знание, пусть даже и в небольшом объеме, изучаемого языка, что, как очевидно, недопустимо ни по каким методическим нормам.

4) С принципом естественности смыкается принцип *вариативности*. Система языка позволяет решать одну и ту же или близкие коммуникативные задачи множеством способов через реализацию в речи различных языковых структур. Выбор того или иного варианта может быть обусловлен экстралингвистическими факторами, что должно найти отражение в процессе обучения.

5) Автоматизация обучения позволяет повысить уровень его мотивации, заинтересованности обучаемого в процессе обу-

чения. Однако такой повышенный уровень мотивации может быть достигнут только благодаря *коммуникативности* обучающей программы (т. е. путем организации диалога обучаемого с машиной) и ее *ситуативной ориентированностью* (т. е. через погружение в ситуацию задания).

При проектировании и разработке АОС приходится решать следующие основные задачи:

— отбор материала и его классификация (распределение по урокам/темам);

— разработка сценария диалога;

— внешнее оформление, или дизайн, программы;

— организация взаимодействия в системе (реализация комплексного подхода к обучению);

— разработка языка представления данных;

— организация баз данных по предметной области и интерфейса для их ведения.

Отношение каждой такой подзадачи к различным типам обучающих программ неодинаково.

Функциональным ядром АОС являются именно обучающие программы, так как собственно автоматизированное обучение имеет место в процессе взаимодействия обучаемого с обучающей программой. Именно это взаимодействие определяет повышенную мотивацию и эффективность обучающих воздействий.

Кроме того, необходимо заметить, что разработка всех других компонентов системы хотя и может быть ориентирована на задачи обучения, все же связана в первую очередь не с обучающей деятельностью: так, база знаний может быть разработана вне всякой связи с ее будущим дидактическим использованием; она может применяться и в исследовательских целях, и для решения иных прикладных задач.

Обучающие программы могут быть классифицированы по множеству разнообразных оснований. Наиболее существенными мы будем считать следующие:

1) допускает ли программа непосредственное взаимодействие обучаемого с ЭВМ. Существует обширный класс программ, которые являются обучающими в широком смысле, однако не выполняют функции собственно обучения. Такие программы носят вспомогательный характер, обеспечивая статистическую или иную обработку информации о ходе учебного процесса, позволяя переложить все «административные» операции (такие, как составление расписания занятий, регистрация и хранение результатов действий обучаемых и обучающих, в том числе программ, составление отчетов, поглощающее много времени у преподавателей, и т. п.) на вычислительную технику;

2) какие из функций, необходимых для реализации законченного фрагмента обучения (контроль, демонстрация требуемой деятельности, оценивание деятельности обучаемого и т. п.), выполняет программа. Очевидно, что алгоритмы для решения

каждой из этих задач обладают чрезвычайно ярко выраженным своеобразием и не могут быть реализованы через один и тот же набор алгоритмических процедур; общими в различных обучающих программах могут быть только некоторые чисто формальные процедуры, не имеющие отношения к процессу обучения вообще (например, оконный интерфейс, процедуры обмена данными с накопителями, заставки и иные средства оформления программы, средства организации диалога с обучаемым и т. п.);

3) каков тип предъявления учебной задачи: обычное; погружение в задание; погружение в ситуацию задания;

4) *допускается ли в программе диалог и какова его форма.* Очевидно, что диалог является достаточно важным средством повышения мотивации обучения; однако он обладает и еще одной чрезвычайно существенной функцией: это — единственное средство осуществления обратной связи «программа — обучаемый». В свете этого не вызывает сомнения, что сколь-нибудь сложные программы без диалога в той или иной форме обходиться не могут. *Без диалога невозможно управление ходом* выполнения программы со стороны обучаемого, а также учет его индивидуальных особенностей. Можно сказать, что при работе с обучающей программой, в которой не предусмотрен диалог, обучаемый вынужден быть абсолютно пассивным.

Диалог может строиться как на специально сконструированном фрагменте какого-либо естественного языка (выбирается, *как правило, либо родной язык, либо английский*), так и на основе искусственной системы команд. Очевидно, что использование естественного языка в обучающих программах очень заманчиво со всех точек зрения, однако сопряжено с большими трудностями, переходя в ранг задач искусственного интеллекта; в то же время специальный язык команд требует предварительного изучения обучаемым и имеет смысл только в случае, когда применяется этот язык не в одной и не в двух обучающих программах, а в обучающей системе;

5) возможно ли управление ходом выполнения программы со стороны обучаемого. Речь идет о линейных и ветвящихся алгоритмах. В первом случае сценарий программы не допускает никаких отклонений от заранее заданной последовательности обучающих, контролирующих и подобных им воздействий, во втором случае их последовательность может произвольно варьироваться по желанию обучаемого. Это повышает активность участия обучаемого в учебном процессе, что, бесспорно, связано, с одной стороны, с повышенным уровнем мотивации обучения, а с другой — с настройкой на уровень обучаемого;

6) учитываются ли индивидуальные особенности обучаемого. Под индивидуальными особенностями понимается не качественное своеобразие, а некая количественная характеристика обучаемого, так называемый уровень обучаемого. Число раз-

личающихся уровней может быть довольно большим (в некоторых действующих АОС — до девяти). Учет индивидуальных особенностей необходим для того, чтобы подход к обучению был индивидуальным, что позволяло бы существенно повысить эффективность учебного процесса в целом. Этот учет может осуществляться как заранее, до начала выполнения программы (в этом случае могут существовать несколько различных программ и оболочка, позволяющая выбрать одну из них), так и в процессе выполнения программы. В последнем случае имеется возможность переходить с одного уровня на другой. Этот переход может осуществляться как по желанию самого обучаемого (если он имеет возможность управлять ходом выполнения программы), который выбирает более простые или более сложные задания, так и автоматически, через анализ по ходу программы успешности предшествующих действий. Последняя возможность представляется наилучшей с точки зрения исключения субъективности в оценке уровня;

7) возможно ли обращение обучаемого к базе знаний. Разумеется, эта возможность характеризует только программы, взаимодействующие с ней, т. е. информирующие, собственно обучающие и тренировочные программы, и является не приемлемой для контролирующих;

8) регистрируются ли результаты выполнения программы в АОС. Эта возможность помимо прочих достоинств способна обеспечить преемственность обучающих воздействий между сеансами работы конкретного обучаемого с АОС. В частности, может регистрироваться сам факт выполнения той или иной программы тем или иным обучаемым, что дает возможность избежать необоснованных пропусков и повторов в учебном курсе. Кроме того, фиксация результатов способна обеспечить настройку на уровень обучаемого;

9) содержится ли дидактический материал в тексте программы или хранится в наборах данных вне программы, являясь, тем самым, не зависимым от нее. В этом случае существует возможность использовать одну и ту же программу с разными наборами данных и таким образом экономить человеческий труд, затрачиваемый на программирование обучающих систем, обеспечивая наряду с этим эффективное и быстрое внесение корректив в материал обучения. Исключительно важное значение приобретает в связи с этим разработка языка представления данных, обладающего гибкостью, достаточной для решения наибольшего числа различных задач АОС.

Литература

Автоматизированные обучающие системы на базе ЭВМ / Под ред. А. Ф. Чернявского. Минск, 1980.

Носенко Э. Л. ЭВМ в обучении иностранным языкам в вузе. М., 1988. 103 с.

Пиотровский Р. Г. Компьютеризация преподавания языка. Л., 1988.

Рыбаков Ф. И., Соколова И. Н., Лоскутникова Л. В. Программированное обучение с использованием ЭВМ. М., 1968.

Савельев А. Я., Новиков В. А., Лобанов Ю. И. Подготовка информации для автоматизированных обучающих систем. М., 1986. 174 с.

Фадеев С. В. ЭВМ в преподавании русского языка как иностранного. М., 1990. 81 с.

СПЕЦИАЛЬНЫЙ ТЕКСТ КАК ПРЕДМЕТ ПРИКЛАДНОГО ЯЗЫКОЗНАНИЯ[1]

Текст (устный или письменный) — основной объект любого филологического исследования. Специальный текст — это текст, главным содержанием которого являются те или иные теории, факты, сведения, рекомендации отдельных наук и отраслей знания. Сюда, прежде всего, относятся научные, технические и деловые тексты. Таковы, например, монографии, статьи по медицине, биологии, геологии, химии, кибернетике, машиностроению, навигации, картографии, учебники для вузов, технические документы, пособия, руководства, чертежи, планы, патенты, лицензии, организационно-распорядительная документация.

Научные, технические и деловые тексты начали исследоваться лингвистами сравнительно недавно, с 30—40-х годов XX в. Сегодня язык науки стал одним из основных полноправных и самостоятельных объектов исследования наряду с языком художественной литературы, литературной разговорной речью и традиционными диалектами. Большой и дискуссионный общефилологический вопрос об отношении языка науки к литературному языку, к языку национальному не является предметом прикладного языкознания.

В последнее время интерес к научному тексту, к языкам науки и техники обострился в связи с теми новыми задачами, которые поставлены перед прикладным языкознанием современной научно-технической революцией (автоматическая обработка текстов на естественном языке, стандартизация терминологии, научно-технический перевод, создание терминологических словарей и банков данных, лингвистическое обеспечение автоматизированных систем). В основе решения различных прикладных задач лежит многоаспектный лингвистический анализ

[1] В главе рассматриваются преимущественно общие и лексико-семантические аспекты специального текста.

научного текста как совокупности разных текстов данной отрасли знания.

Специальный текст всегда представляет, репрезентирует то или иное научное, техническое или организационно-деловое знание. Рассмотрим процесс формирования специального текста на примере научных текстов. Научное знание (включая сюда и знания технические) — это совокупность идеальных образов в сознании человека, отражающих явления, свойства, отношения и законы материального мира в сфере науки и техники.

Однако специальное научное знание бесспорно выступает как основной фактор формирования научного текста. Оно многозвенно, многоступенчато, ассоциативно, начиная от общих сведений о той или иной отрасли знания и кончая глубинными концепциями, связанными с отдельными узкими направлениями. Например, филолог обладает определенным достаточным уровнем общефилологических знаний в области языка и литературы и одновременно он, как правило, является специалистом в той или иной относительно замкнутой области — в словообразовании, фонетике, фокльклористике, текстологии и т. п.

Основным внутренним мыслительным фактором порождения нового научного текста является познание непознанного. Подобное движущее творческое начало имеет сложный психофизиологический характер и, конечно, не может быть сведено только к модели научного знания. Но в то же время именно общественный уровень профессионального знания той или иной проблемы является первопричиной, ведущей к порождению научного текста. Всякий исследователь стремится к познанию нового, к открытию terra incognita в своей области.

Научное знание делится на две основные части: знание общенаучное и знание собственно профессиональное, связанное с данной наукой. Общенаучное знание — это знание об основных *наиболее общих понятиях, используемых в разных науках.* Это такие понятия, как метод, система, структура, анализ, синтез, иерархия. Общенаучное знание основных объектов свойств и методов функционирует *по отношению к знанию собственно* профессиональному как управляющая система более общего порядка, что и находит свое отражение в языке в общенаучной и межотраслевой лексике.

Не последнюю роль в формировании научного текста играет и обыденное, «обиходное» знание о мире. Оно все время выступает как своеобразный фон, шкала сравнения степени специализации научного текста. Роль обиходно-бытового знания особенно ярко видна на текстах XVIII — начала XIX в. В XVIII в. в естествознании и географии еще не было сложившейся системы научных знаний, и первые русские академики-путешественники И. И. Лепехин, Н. Я. Озерецковский, И. Георги, С. Г. Гмелин, П. С. Паллас, В. Ф. Зуев строили тексты

своих описаний, опираясь в основном на знание обиходное, на общенародную и местную диалектную лексику.

По отношению к зрелому научному знанию особо выделяется этап знания донаучного. Для некоторых наук, таких как биология, география, история, физика, изучение этого периода развития знания, когда еще только закладываются основы представления об отдельных предметах, явлениях и процессах, существенных для той или иной науки, играет большую роль. Например, историю формирования научных биологических текстов по фауне и флоре нельзя понять вне многовекового донаучного периода накопления народных знаний и представлений о мире рыб, зверей, птиц и растений. При этом изучение проблемы становления структуры текстов требует сведений об уровне знаний и картине видения мира не только у народа, говорящего на данном языке, но и у многих других народов, с которыми этот народ контактировал в разное время. Так, например, русская научная ботаническая и зоологическая терминология в XVIII—XIX вв. во многом сложилась под влиянием тех знаний, которыми обладали ранее финно-угорские и тюркские народы.

Научный текст и как совокупность текстов разных авторов, и как отдельный индивидуальный текст, и как отдельная теория не существует вне истории науки, вне эволюции научного знания. Именно в связи с этим следует особо подчеркнуть значение изучения истории формирования научного текста. Процесс формирования письменного научного текста как совокупности текстов разных авторов может быть достаточно полно раскрыт только через изучение функционирования этого типа текста в течение ряда отдельно взятых хронологических периодов, синхронных срезов, путем моделирования морфологии, лексики, семантики, синтаксиса этих текстов в тесной связи с историей данной науки. Историю языка русской науки нельзя понять и вне истории Российской Академии наук. История Российской Академии наук — это живая история русской науки, воплощенная в ее языке. История языка русской науки немыслима без анализа языка произведений М. В. Ломоносова, П. С. Палласа, И. И. Лепехина, Г. Георги, В. Ф. Зуева, В. С. Якоби, Э. Х. Ленца, Н. Н. Зинина, К. М. Бэра, К. Ф. Рулье, А. М. Бутлерова, И. М. Сеченова, И. П. Павлова, С. А. Чаплыгина, П. П. Семенова-Тян-Шанского, В. Л. Комарова, Л. С. Берга, классиков русской истории, философии и филологии.

В ходе своей эволюции научное знание, достигнув определенной стадии зрелости, ищет знаковую систему, язык своего выражения. И здесь различают две основные формы представления научного знания: 1) сам код, непосредственный язык для записи содержания знания в виде символов, словосочетаний, предложений, знаков их связи и т. д.; 2) различные типы

текстов, в которых на том или ином языке находит свое оформление научное знание. При этом тексты вторичны по отношению к коду и существуют только при наличии языка для записи знания. В свою очередь, сами типы текстов во многом определяются традицией, историей науки и книжного дела.

Разработка и поиск знаков для описания знаний — это первый этап оформления и представления научного знания, первый этап превращения его в научную информацию. Никакая информация не существует вне ее кодового представления, вне текста. На поверхностном уровне в качестве такого кода всегда выступает та или иная система знаков как язык. Функции такого языка может выполнять один из классических мертвых языков (древнегреческий, латинский) или искусственный язык.

Язык науки чаще всего представляет собой сочетание искусственного и естественного языков, при этом первый обычно функционирует в сфере терминологии и обозначения узких понятий, а второму отводится роль языкового оператора и диспетчера (предикаты, связки, категоризаторы, модели морфологии и синтаксиса). Таково большинство современных языков науки и техники (подъязыки биологии, медицины, математики, физики, геологии).

На этапе кодирования научное знание сталкивается с совсем другой системой — с системой знаний о языке, заложенной в мозге человека. Языковая система функционирует как множество, включающее в себя подмножество различных систем. Во-первых, это собственно языковые системы — такие, как фонологическая, морфологическая, синтаксическая, лексико-семантическая. Эти системы отражают разные уровни внутреннего строения языка. Во-вторых, это целый ряд функциональных систем и подсистем, таких, как системы диалектного языка, языков науки и техники, система стандартного литературного языка.

В свою очередь, многие из этих систем распадаются на ряд подсистем. Так, например, морфологическая система делится на подсистемы морфемики, словообразования, формообразования. Система диалектного языка делится на подсистемы отдельных диалектов, системы языка науки и техники — на подсистемы языков отдельных наук и отраслей знаний (подъязыки физики, биологии, электроники, машиностроения). Система стандартного литературного языка делится на подсистемы языка художественной литературы, языка газет, языка радио и телевидения, подсистему разговорной речи.

Каждую систему образует множество объектов-понятий и разных типов отношений между ними.

Таким образом, система языка представляет собой тип сложной иерархической системы. При этом семантика не образует отдельного уровня. Напротив, семантические отношения

пронизывают все уровни языка, за исключением фонологического.

Постепенно содержание научного знания начинает проникать и в знаки избранного нами языка, насыщать и заполнять их. В языке слово, словосочетание уже неотделимы от их значения, и здесь содержание научного знания становится элементом языка науки. Научное знание, нашедшее свое выражение в слове, в термине, переходит в качественно новую стадию, включаясь в семантическую систему и структуру того или иного языка науки, становясь компонентом лексико-семантической системы этого языка.

Термин (включая научно-технические термины и термины организационно-распорядительной документации) — это единица какого-либо конкретного естественного или искусственного языка (слово, словосочетание, аббревиатура, символ, сочетание слова и букв-символов, сочетание слова и цифр-символов), обладающая в результате стихийно сложившейся или особой сознательной коллективной договоренности специальным терминологическим значением, которое может быть выражено либо в словесной форме, либо в том или ином формализованном виде и достаточно точно и полно отражает основные, существенные на данном уровне развития науки и техники признаки соответствующего понятия. Термин — слово, обязательно соотносимое с определенной единицей соответствующей логико-понятийной системы в плане содержания.

Глубинный признак терминов позволяет отделить их от других единиц языка и расчленить все множество терминов. Этим глубинным признаком терминов является обозначение ими общих понятий. Поскольку существует несколько типов общих понятий, могут быть выявлены и разные типы терминов[2].

Прежде всего выделяются самые общие понятия материи и ее атрибутов, которые носят названия категорий (*материя*, *пространство*, *время*, *количество*, *качество*, *мера* и др.). Соответственно термины, которые обозначают категории, представляют собой тип терминов категорий.

Далее, на каждом этапе развития человеческого знания, в каждую эпоху фигурирует определенное количество общенаучных и общетехнических понятий, которые используются в любой науке (отрасли техники) (*система*, *структура*, *метод*, *закон* в науке, *надежность* в технике). К ним примыкают общие понятия методологических наук — философии, общей теории систем, кибернетики, информатики и др.; некоторая часть понятий этих наук может использоваться, как и общенаучные понятия, в различных областях знания (например, *информация*, *элемент*). При этом нужно иметь в виду, что общенаучные (об-

---

[2] Приводимая здесь классификация терминов разработана В. М. Лейчиком.

щетехнические) и межотраслевые понятия являются таковыми не потому, что они используются в ряде отраслей знания, а потому, что они имеют общее содержание, которое и позволяет применять их в разных отраслях, в большинстве случаев добавляя к общему содержанию конкретные признаки. В качестве примера можно привести общенаучный термин *метод*, межотраслевой термин *аналитический метод* и конкретно-научные термины *математический метод*, *метод Монте-Карло* (ср. также межотраслевой термин *кластер* — малый коллектив частиц, от англ. cluster — груда, скопление, пучок, гроздь, и его применения в физике, химии, астрономии, биологии, социологии, общей теории систем, а также в науковедении и информатике).

Наконец, в каждой области знания и деятельности имеются специфические понятия разной степени обобщенности: от самых крупных — классов (родов) до наименьших — видов, а также понятий, которые отражают аспекты рассмотрения этих классов. Эти два типа понятий названы видовыми и аспектными: в качестве примеров могут быть приведены понятия геологии: *субдельта* (видовое понятие относительно понятия *дельта*), *геохимия нефти*, *запасы нефти* (аспектные понятия).

Все вышеперечисленные типы понятий находят свое языковое воплощение в типах терминов. Выделяются термины категорий, общенаучные и общетехнические термины, межотраслевые термины, специальные термины (примеры были приведены при перечислении типов понятий).

Известно, что типология является основой классификации. В этом смысле описанная здесь типология терминов — членение терминов по наиболее важным их признакам — представляет собой собственно терминоведческую классификацию терминов. Основанием же всех последующих классификаций служат различные отдельные признаки терминов — содержательные, формальные, функциональные, внутри- и внеязыковые. Все эти классификации могут быть связаны с теми науками и областями знания, в которых они используются.

Первой классификацией терминов по содержанию, используемой преимущественно в философии, является деление на термины наблюдения и теоретические термины. За терминами наблюдения стоят классы реальных объектов, а за теоретическими терминами — абстрактные понятия, зависящие обычно от определенной теории, концепции. Такое деление является достаточным для решения терминологических проблем философии (философии науки), но для решения философских проблем терминоведения приходится построить более дробную классификацию, поскольку степень абстрактности понятий, обозначаемых теоретическими терминами, различна: от философских категорий до общенаучных и специальных научных понятий. Так, в систематике животных выше таксонов лежат так называемые таксономические категории (неконкретные объ-

екты) — виды, подсемейства, классы. Важно, что эта иерархия может находить выражение и в формальной структуре терминов. В частности, в той же системе К. Линнея наименования таксонов (термины наблюдения) включают в себя наименования таксономических категорий: Betula pubescens — *береза пушистая.*

Второй классификацией терминов по содержанию — по объекту названия — является распределение их по областям знания или деятельности, или, иначе говоря, по специальным сферам. Перечень этих сфер может быть обобщенно представлен следующим образом: наука, техника, производство; экономический базис; надстройка. Основываясь на этой социологической схеме, можно сформулировать перечень рубрик, входящих в классификацию терминов по области знания.

В сфере науки выявляется группа научных терминов. Она распадается, говоря в общем, на столько классов, сколько имеется наук на определенном этапе научно-технического прогресса; а в каждом классе физических, химических и других терминов выделяется столько группировок (терминосистем), сколько существует различных независимых теорий описания физических, химических и иных объектов и закономерностей. Что касается различий так называемой научно-технической и общественно-политической терминологии, то, прежде всего, политические науки (теория государства и права, международные отношения и др.), по единодушному мнению специалистов, входят в число общественных наук, и следовательно, политические термины включаются в множество общественных терминов. Далее, все эти термины обозначают научные понятия в той же мере, что и так называемые научно-технические термины; разница заключается только в том, что первые обозначают понятия общественных наук, а вторые — понятия естественных и технических наук. Поэтому если стремиться к точности, то целесообразно говорить о терминах общественных, естественных и технических наук и о технических терминах и терминологиях, а не о научно-технической и общественно-политической терминологии. Однако термины общественных наук обладают рядом специфических признаков, которые противопоставляют их терминам естественных и технических наук. Это: 1) прямая, явно выраженная зависимость терминов общественных наук от определенной теории, определенной системы взглядов. При ближайшем рассмотрении термины естественных и технических наук также зависят от теории, которая, в свою очередь, определяется мировоззрением (например, *параллельность* в геометрии, *масса* в физике), но эта зависимость может быть затушевана. В терминах социальных наук она входит в их содержательную структуру; 2) своеобразная реализация признака системности. Наряду со стройными терминосистемами, отражающими законченные теории (политическая экономия, философская система

Гегеля), имеются такие области знания, для которых не построены системы *понятий и терминосистемы* (*например, описание танца, мода* и др.); 3) наличие терминов с размытыми границами обозначаемых ими понятий, например, терминов, обозначающих общие понятия социального характера (*личность, идеал*); 4) более широкое развитие синонимии и многозначности, чем в терминосистемах естественных и технических наук (*язык* — многозначный термин); 5) включение оценочного фактора в семантику терминов (ср. термины типа *ренегат, красные, белые*, а также *зеленые* 'партии зеленых').

В сфере производства и техники функционируют технические термины. Это единицы языка, обозначающие машины, механизмы, инструменты, операции. Технические термины отличаются от научных терминов прежде всего меньшей зависимостью от *концепций использующих их людей, хотя такая* зависимость и существует. В наше время технические термины нередко проникают и в собственно научные издания.

В сфере экономического базиса и производственных отношений имеются, с одной стороны, термины языка описания (в языках политической экономии, конкретной экономии) и, с другой стороны, термины языка обслуживания экономики. Лексическими единицами языка обслуживания служат такие сочетания терминов, как *ввод в действие общей (полезной) площади жилых домов,* которые называются показателями. Показатели — это совокупность признаков, характеризуемых приводимыми данными. В наименование показателя входят термины, обозначающие: а) характеризуемый (измеряемый) объект экономики (*продукция, работающие*); *б) состояния, свойства* этих объектов и процессы, которые с ними совершаются (*наличие* или *численность* (работающих), *производство* (продукции)); в) формальный способ (алгоритм) вычисления показателя, например, *объем* (реализации продукции).

В рамках надстройки выделяются административно-политическая сфера (куда входят оборона, юстиция, внешние сношения и др.) и социально-культурная сфера (здравоохранение, наука, культура, образование и др.).

В административно-политической сфере выделяются, в частности, термины языка управления, в том числе термины делопроизводства, языка дипломатии, военных языков. Правда, часть военных терминов сближается с техническими, другая — с *экономическими, третья — с научными терминами.* Военные термины — типичный пример терминов стыковой области. На их примере можно показать, что вообще граница между терминами, выделяемыми по области знания, достаточно зыбка. Так, многие термины технических наук могут одновременно являться техническими терминами (*прокатка, волочение, отжиг* и т. п.), а многие термины документоведения фигурируют и в терминологии делопроизводства (*акт, архив*). Тем не менее клас-

сификация терминов по объекту называния чрезвычайно важна: она отражает уровень науки и развитости общественной структуры на определенном этапе.

К социально-культурной сфере кроме терминов общественных наук (политэкономии, социологии, этнографии) относится так называемая общественно-политическая лексика. Строгое различение терминов (общественных наук) и общественно-политической лексики важно, в частности, для построения информационных языков.

Классификация терминов по объекту называния внутри отдельных областей знания является наиболее подробной классификацией терминов.

Третья содержательная классификация терминов — по логической категории того понятия, которое обозначается термином. Выделяются термины предметов (*млекопитающие*), процессов (*умножение, делопроизводство, компрессия*); признаков, свойств (*хладноломкость*), величин и их единиц (*сила тока, ампер*).

Лингвистические классификации терминов основаны на признаках терминов как слов или словосочетаний определенного языка.

Классификация по содержательной (семантической) структуре позволяет выделить однозначные термины (*шунтирование, гайка, хромосома*) и многозначные термины, т. е. такие, которые имеют два или более значений в рамках одной терминосистемы (*суд* — 1. совокупность судей и заседателей; 2. судебное заседание; 3. здание суда)[3]. С точки зрения семантики выделяются термины — свободные словосочетания (*муфельная печь, справка с места жительства*) и устойчивые (в том числе фразеологические) словосочетания (*всемирное тяготение*).

Классификация терминов по формальной структуре является очень дробной. Прежде всего, выделяются термины-слова. Они, в свою очередь, подразделяются на корневые (*вода*), производные (*предлог, делитель, пересортица*), сложные (*обществоведение, биосфера*), сложносокращенные (*капвложения*), а также слова необычной структуры — телескопические (*магнитола* — от *магнитофон + радиола*), с обратным порядком звуков (*мо* — от *ом*), цепочечные образования (*синтез-газ, 2,5-диметил-5-этил-3-изопропилгептан, система хребет — хребет — хребет*).

Далее, выделяются термины-словосочетания. Наиболее распространенными структурами здесь являются сочетания существительного с прилагательным, существительного с существительным в косвенном падеже (*степень свободы*), существительного с другим существительным в качестве приложения (*швея-мотористка*). Имеются также многословные термины, иногда со-

---

[3] Термины, имеющие одинаковую форму и входящие в разные терминосистемы, являются омонимами.

стоящие более чем из 5 слов (*фильтрационный потенциал самопроизвольной поляризации в скважине* — термин ГОСТа).

Характерными явлениями в формальной структуре терминов являются усечение однословных терминов (*кино* — от *кинофильм* или *кинотеатр*) и сокращение (аббревиация) многословных терминов. Наблюдается множество видов аббревиатур: буквенные (*к. п. д.*), звуковые (*ЖЭК*), слоговые (*горком*), словоподобные (*сигран* — от *синтетический гранит*), полностью совпадающие со словом (*ГАЗ* от: *Горьковский автомобильный завод*); кроме того, сочетания аббревиатур со словами (*МГД-генератор* — от *магнитогидродинамический генератор*).

Постоянно появляются термины специфической формальной структуры с использованием элементов искусственных языков; символы-слова (*α-частица*), модели-слова (*i-балка*, т. е. двутавровая балка по сходству с буквой *I*).

Классификация по мотивированности / немотивированности показывает, что есть термины, значение которых может быть или не может быть объяснено их структурой. Здесь различают термины, полностью мотивированные (*газопровод*), частично мотивированные (*болезнь Паркинсона*), полностью немотивированные (*ромб*), а также ложномотивированные (*громоотвод*).

В зависимости от языка-источника различаются термины исконные (*датчик*), заимствованные (*дисплей* — англ., *квершлаг* — нем.), гибридные (*металловедение*, *антиобледенение*).

С точки зрения принадлежности терминов к частям речи различают термины-существительные, прилагательные, глаголы, наречия. Например, среди лингвистических терминов есть существительные (*залог*, *вид*), прилагательные (*немотивированный*, *парасинтетический*, *сочинительный*). Среди терминов музыковедения фигурируют термины-наречия (*пиано*, *пианиссимо*). Подсчеты показывают, что терминов — названий объектов в процентном отношении гораздо больше, чем терминов — названий признаков. Да и обозначения признаков в терминах часто выступают в опредмеченном виде.

Классификация терминов по авторству отражает социологический подход к терминам. Известны в этом плане коллективные и индивидуальные термины. Так, термин *геликоптер* создал Леонардо да Винчи, термин *промышленность* — Н. М. Карамзин, термин *социология* — О. Конт, термин *холодная война* — У. Липпман.

По сфере использования выделяются универсальные (для многих родственных областей), уникальные (для одной области) и концепциально-авторские термины; например, лингвистические термины могут обозначать явления, характерные для всех языков (*фонетика*), для одного или нескольких языков (*эргативность*) или только для одного подхода (*глоссематика* — термин Л. Ельмслева).

Важная роль терминов в процессе научного познания объек-

тивной действительности позволяет построить их науковедческие классификации. Так, выделяются термины, служащие для фиксации знания, термины, используемые как инструмент познания, и термины обучения. Например, термином — инструментом познания следует назвать первоначально введенный В. Рентгеном термин *х-лучи*, впоследствии замененный в русском языке термином *рентгеновское излучение*. Термины обучения используются в начальной школе в связи с их простотой и понятностью (*правописание*), затем их заменяют научными терминами — средствами фиксации знания (*орфография*).

Для каждой эпохи создается историко-лексикологическая классификация терминов, в которой фигурируют термины-архаизмы, термины-неологизмы. Эта классификация теснейшим образом связана с вышеприведенной классификацией терминов по объекту называния. Неологизмов больше в формирующихся терминосистемах. Архаизмы характерны для концепций тех терминосистем, которые уходят в прошлое в связи с ростом научного знания и выяснением устарелости некоторых научных воззрений. Однако поскольку термины остаются в языке как его лексические единицы (хотя и детерминологизированные), они могут возродиться в составе новых терминосистем или в новом значении при развитии терминосистемы. Так, в последние годы возродились термины *наставник, аэробус* и др.

В связи с тем, что термины выполняют прикладную функцию как инструменты познания и как средства фиксации научного или технического знания, они подвергаются унификации и закрепляются в той или иной форме в качестве рекомендованных или стандартизованных. На этой основе строится классификация терминов по нормативности — ненормативности, которая включает в себя термины, находящиеся в процессе стандартизации (стандартизируемые), подвергшиеся стандартизации (стандартизированные) (*насос*), отклоняемые в процессе стандартизации (недопустимые) (ндп. *помпа*); находящиеся в процессе упорядочения (рекомендуемые), подвергшиеся упорядочению (рекомендованные) (*энергия Гельмгольца* — в термодинамической системе), параллельно допустимые (*изохорно-изотермический потенциал* — там же, в том же значении), отклоненные в процессе упорядочения (прк. *свободная энергия* — то же) (Сборник..., 1973, с. 18). К этому следует добавить, что в сфере науки и техники существуют термины, нормативность которых является обязательной: так, термины радиочастот нормализованы в международном масштабе, поскольку это необходимо для обеспечения безопасности мореплавателей, авиаторов и др. (Как правило, эти термины являются интернациональными; по крайней мере, их семантика подвергается нормализации решениями международных организаций.)

Наконец, в результате анализа частоты использования терминов в текстах может быть применена классификация, выде-

ляющая высокочастотные и низкочастотные термины. Сведения о частотности терминов могут быть почерпнуты из многочисленных частотных терминологических словарей, изданных в нашей стране.

Приведенный перечень *классификаций терминов позволяет* сделать вывод о том, что столь многогранное явление, как термин, входит в самые разные классификации — по логическому, лингвистическому, науковедческому и другим принципам. Эти классификации в своей совокупности характеризуют роль и место терминов в научной, экономической, политической, управленческой и других сферах функционирования современного общества. Но термины — всегда часть текста.

Кодирование научного знания всегда связано с созданием какого-то текста. Обретая язык, на следующем этапе научное знание ищет разные формы самовыражения и находит его в разных видах первичных устных и письменных текстов (доклад, диссертация, споры о терминах, статья, монография и т. д.). Текст — следующая качественно новая форма существования научного знания. Здесь содержание знания функционирует уже как семантика текста, а само знание постепенно достигает статуса научной информации. Знание становится научной информацией только в системе общественных научных коммуникаций.

При этом и сам текст выступает уже как одна из форм представления семантики научного знания.

Устными формами представления научного знания являются дискуссия, диспут, обсуждение проблемы в узком кругу специалистов, когда не только уточняется содержание понятий, но иногда получается и новое знание.

Наиболее полно знание об объекте, его признаках и свойствах отражается в жанре научной монографии (книга, диссертация), даже и в том случае, если она представляет лишь одну из школ, одно из мнений. В серьезной монографии всегда найдется место для обсуждения чужих идей. Именно в монографии широкого профиля нередко всесторонне отражаются и междисциплинарные связи.

К научной монографии примыкают учебники для вузов. Их выгодно отличает наличие определений, четкость построения и изложения.

Статьи, как правило, обсуждают один из аспектов, одну из сторон, связанных с тем или иным научным понятием, с той или иной проблемой.

Таким образом, и доклад, и дискуссия, и монография, и учебник, и статья — все эти типы текстов в разной степени демонстрируют научное знание, его семантику.

По-своему, в упрощенной форме представляют научное и техническое знание такие документы, как технические руководства по эксплуатации и ремонту машин, паспорта товаров, технические проекты и описания станков, изделий.

Различные формы репрезентации семантики научного знания отражают разную, заранее заданную меру глубины его представления.

Достигнув определенного, достаточно высокого уровня развития, все более обретая свойства информации, научное знание находит новые каналы коммуникативных связей в виде различных компрессированных текстов. Особую группу таких текстов, представляющих научное знание, составляют вторичные документы (тексты). Это словари, энциклопедии, классификаторы, рубрикаторы, стандарты на термины, нормативные справочники, рефераты, аннотации, популярные лекции, беседы, интервью, реклама. Все эти типы текстов передают научное знание уже в компрессированном виде и, исходя из своей прагматики, стремятся сосредоточить внимание на самом главном, на основном. Общенаучная тенденция к компрессии знания, к превращению его в удобный тип научной информации — это то, что характеризует все вторичные тексты. При этом такие жанры, как энциклопедия, отраслевой энциклопедический и толковый терминологические словари, отличаются, как правило, большой глубиной отражения самого знания.

Структура текстов формирует и кристаллизует научное знание. Чем стройнее типовая архитектоника текста, тем, как правило, строже язык описания понятий, употребление терминов. Одновременно на язык науки во всех типах текстов влияет и языковая норма как таковая. Именно норма выступает в качестве того фильтра, который отделяет то, как можно сказать, от того — как нельзя.

Тексты открывают путь к новым, более высоким ступеням знакового представления научного знания, а именно к моделированию логико-понятийных систем и систем собственно терминологических (терминосистем).

Выявление логико-понятийной структуры текста предусматривает, прежде всего, тщательно поставленный эмпирический этап. На этом этапе исследователь ставит вопрос о моноаспектности или полиаспектности данной науки, отрасли знания. В случае полиаспектности выявляются основные направления, устанавливаются связи и взаимоотношения, с одной стороны, отдельных аспектов между собой, а с другой — отдельных научных направлений с рядом соседних и сопредельных наук. Именно на этом этапе исследователь должен внимательно ознакомиться с основными моментами в эволюции данной науки, с тем, как происходит процесс ее дифференциации.

На стадии эмпирического анализа исследователь логико-понятийной структуры текста определяет иерархию и взаимодействие выделенных аспектов и направлений. При этом он может и не быть узким специалистом в данной области, исследователем может быть и лингвист, но он всегда должен достаточно разбираться в общем состоянии данной науки и работать в не-

посредственном контакте с некоторыми крупными и широко эрудированными специалистами в данной отрасли. Шаг за шагом, идя от общего к частному, он устанавливает систему общих иерархических отношений, характерных для данной отрасли знания.

Этап выделения основных логико-понятийных отношений имеет весьма существенное значение, так как именно в результате выявления общих иерархических отношений могут быть выделены такие классы, компоненты, которые могут рассматриваться и как вполне самостоятельные аспекты направления данной отрасли знания. Само моделирование терминосистемы целесообразно вести поаспектно, по отраслям, направлениям, в соответствии с этапами истории науки, отрасли знания.

Довольно сложен вопрос о глубине такой модели. Чаще всего глубина определяется целью исследователя, задачей. Могут быть довольно глубокие модели, ориентированные на создание отраслевой энциклопедии, терминологического словаря, лингвистическое обеспечение конкретных информационно-поисковых систем. Например, создание словарей названий пресноводных рыб СССР и промысловых рыб мировой фауны потребовало в свое время предварительной разработки иерархической системы всех видов промысловых рыб мира, а небольшая сама по себе задача создания информационного словаря-тезауруса по техническим средствам навигации заняла около двух лет у филологов и специалистов отрасли.

Содержание современного научного знания в целом едино для специалистов из разных стран, и в этом отношении глубоко справедливо мнение о том, что наука как знание неделима по государствам и народам. Так, например, системы представлений о видах животных и растений в биологии, о болезнях в медицине, о типах рельефа в географии, о навигационных приборах в технике в целом в основе своей едины для всех специалистов в данной области знания, говорящих как на английском, так и на русском, немецком или французском языках. Но если содержание научного или научно-технического знания о предмете в принципе едино у специалистов из разных стран, говорящих на разных языках, то модель логико-понятийной системы этого знания может быть представлена как на искусственном языке, имеющем интернациональный характер (система латинских обозначений в биологии, медицине, символов в химии, математике, физике), так и на естественном национальном языке.

Построение логико-понятийных схем, моделей текстов хорошо выявляет многие семиотические черты языков науки. Так, именно логико-понятийная модель ярко демонстрирует тот семиотический барьер, который порой оказывается сильнее языкового барьера. Мы можем одинаково неплохо владеть англий-

ским, французским или немецким языками и не понимать смысла статьи и монографии по судостроению или биохимии, написанных как по-русски, так и на других языках. Мы не понимаем смысла таких текстов не только и не столько потому, что не знаем терминов, а прежде всего потому, что нашему сознанию не свойственна понятийная модель знания данной науки. Этим объясняется, в частности, и то, что переводы с языка на язык узкоспециальных исследований, сделанные крупными учеными в своей области, как правило, по уровню выше переводов, сделанных филологами.

Следующий этап в сближении научного знания с системой языка — вхождение знания в терминосистемы отдельных наук. Терминосистемы чаще всего находят свое выражение уже в форме естественных национальных языков (системы английских, русских терминов по радиофизике, по биологии и т. д.). Здесь, на этапе моделирования терминосистем, научное знание вплотную врастает в систему того или иного конкретного естественного языка.

Термины, слова и словосочетания оформляются в лексико-семантическое поле, выстраиваются в лексико-семантические парадигмы. Если модель логико-понятийной системы научного знания является основным компонентом абстрактного знания и притом нередко в форме искусственного языка, то терминосистема как совокупность слов и словосочетаний, представленная в плане выражения на том или ином естественном национальном языке, выступает уже как фрагмент лексико-семантической системы данного подъязыка.

Терминосистема формируется на базе логико-понятийной системы. При этом отдельные звенья логико-понятийной системы заполняются терминами того или иного национального языка.

Построение терминосистем как фрагмента модели знаний о мире предусматривает вскрытие всех типов связей и взаимоотношений между понятиями. Основные виды таких отношений — это гипонимия (род — вид), соподчинение на одном уровне — парциация (часть — целое), ассоциация (локализация объекта, его назначение, характеристика по квалитативным и квантитативным признакам, по времени, форме, размеру, весу и т. д.). В каждой системе знания любой объект в каком-то отношении нередко является и признаком, а почти каждый признак в другом отношении выступает как объект.

Уже на концептуальном уровне модель знания ищет свой план выражения. И здесь вопрос о форме представления модели знания или ее фрагмента — один из кардинальных в общей стратегии разработки моделей знания о мире. Терминосистема — это одна из возможных, хотя и не единственная форма представления любой такой модели и, по-видимому, форма, которую не может миновать ни одна модель знания в своей внут-

ренней эволюции. В принципе логически можно было бы поставить вопрос так: сначала логико-понятийное моделирование знаний, а потом поиск форм ее выражения. В то же время, как известно, формой любого понятия являются именно знаки языка, и прежде всего языка естественного. И в этом отношении любое моделирование знаний в плане выражения всегда протекает в форме слов, словосочетаний и высказываний. Нет и не может быть моделей знания, оторванных от языка.

Таким образом, моделирование знаний и построение терминосистем — это две стороны одного и того же процесса — разработки модели знания. Любое построение терминосистемы всегда есть моделирование знаний, и любое моделирование знаний невозможно без построения терминосистем.

Каждая модель может быть представлена как в текстовой, так и в табличной форме или в виде графа. Впоследствии такая модель может быть спроецирована на разные тексты с целью насыщения ее реальной терминологией того или иного языка.

Чаще всего используются классификационные сети, в которых в узлах сети помещаются термины — наименования понятий, а дугами отображаются отношения между понятиями. Сети делятся на иерархические и разветвленные.

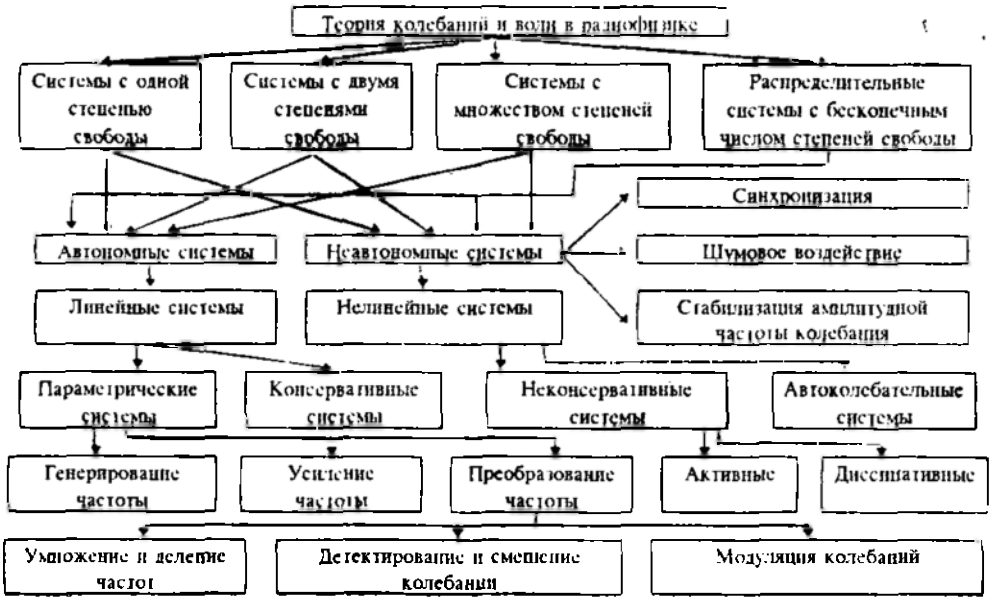
Сетевое представление данных имеет не только чисто прикладное значение, оно позволяет глубже проникнуть в систему логики данной науки, точнее смоделировать терминосистемы отдельных отраслей и направлений. Глубоко разветвленные типы семантических сетей позволяют отражать динамические аспекты структурного научного знания (подробнее см. раздел «Сетевое представление лингвистических данных»).

Приведем пример иерархического сетевого представления данных, разработанных в Нижегородском университете в связи с работой над терминологическим словарем по радиофизике (см. рисунок на с. 84).

Именно в моделях логико-понятийной системы лучше всего предстает многоступенчатость, иерархичность и ассоциативность *научного знания*.

На этапах проникновения и врастания научного знания в формы естественного языка особую роль начинают играть собственно языковые грамматические средства. Подобные элементы (предикаты, наречия, местоимения, служебные слова, морфемы, формы словоизменения) непосредственно в терминосистему не входят, но именно они вместе с морфологической системой языка выступают в качестве динамических процессоров порождения текста в речевой деятельности.

В изучении морфологии языка науки наибольшие результаты дает применение квантитативных методов. Поскольку любое научное произведение или сочинение по технике в письменной форме создается на том или ином литературном языке и


Теория колебаний и волн в радиофизике
Системы с одной степенью свободы
Системы с двумя степенями свободы
Системы с множеством степеней свободы
Распределительные системы с бесконечным числом степеней свободы
Синхронизация
Автономные системы
Неавтономные системы
Шумовое воздействие
Линейные системы
Нелинейные системы
Стабилизация амплитудной частоты колебания
Параметрические системы
Консервативные системы
Неконсервативные системы
Автоколебательные системы
Генерирование частоты
Усиление частоты
Преобразование частоты
Активные
Диссипативные
Умножение и деление частот
Детектирование и смешение колебаний
Модуляция колебаний

впоследствии проходит стадии обсуждения, апробации, литературного редактирования, вычитки, корректуры, то естественно, что оно, как правило, лишено отклонений от норм литературного языка. Во всех научных и технических документах мы находим репрезентацию стандартных, нормативных морфологических парадигм. Все различия здесь лежат в частоте, доле распределения форм и категорий по отдельным типам текстов. В то же время поскольку лексическое и грамматическое в слове тесно связаны, то и лексическое значение слова не существует вне его грамматического оформления. Таким образом, научный текст является результатом контаминации терминосистемы с морфологической и синтаксической системами.

В этом процессе семантика терминосистемы взаимодействует с чисто категориальными грамматическими значениями (значение морфем, частей речи, служебных слов, словосочетаний).

В конечном счете, всякое научное знание воплощается в тексты и познается только через тексты. Любая абстрактная энциклопедическая модель научного знания есть производное от множества реальных текстов и представляет собой репрезентацию этих текстов на уровне семантических моделей. Семантическая модель языка науки, отрасли — это инвариант микромоделей отдельных частных текстов, которые выступают как варианты этого языка.

В чисто лингвистическом плане моделирование семантики научного текста целесообразно вести в терминах теорий семантического поля и компонентного анализа. Логико-понятийная структура текста, интерпретируемая в семантическом плане как единство семантических макро- и микрополей, отражает основные элементы семантической парадигматики текста, но не совпадает со структурой конкретных текстов и книг.

Трудно переоценить роль научного знания и языка в формировании научного текста; в то же время можно выделить целый ряд других компонентов, которые играют не меньшую роль в формировании научного текста и действуют императивно в основном еще на доязыковом уровне. Обычно такие своеобразные императивы порождения текста проявляются в той или иной конкретной коммуникативной ситуации. Это такие компоненты-императивы, как прагматика, коммуникативная среда, традиции текста.

Прагматическая направленность исследования предопределяет свои компоненты научного текста как на уровне макроструктуры, так и на собственно языковом уровне. В содержательном плане сюда относятся определение цели и задач работы, указания на то, кому она предназначена, адресована, на практическое значение проведенного исследования. В плане макроструктуры это прикнижная аннотация и соответствующие фрагменты предисловия, введения. Пространные указания на то,

как пользоваться книгой, находим в учебниках, словарях, энциклопедиях. Фактор прагматики находит самое различное выражение в разных языках. Чаще всего это модальные слова, частицы, наречия, адвербиальные конструкции.

Фактор коммуникативной среды выступает как проявление тех информационных связей и отношений, в которые включен исследователь. В процессе создания специального текста, как правило, обязательно учитывается объем профессиональных знаний, специализация и интересы тех, кому адресуется этот текст. Таким образом, фактор коммуникативной среды выступает одновременно и как императив адресата. Данный императив особенно ярко проявляется в том, что каждый исследователь сознательно, по необходимости стремится писать так, чтобы его поняли все те, кто работает в этой области.

В учебниках для студентов, в научных работах широкого профиля встречаем указания на уровень знаний, необходимых для понимания и усвоения соответствующего материала. Именно через фактор коммуникативной среды в текст незримо проникают и именно через текст реализуются опосредованные неформальные информационные связи ученых между собой. Можно прекрасно представлять себе образ специалиста, живущего в другой стране, никогда с ним не встречаясь, но хорошо зная его работы.

Коммуникативная среда предопределяет наличие различных связей между каждым новым текстом и текстами, ранее созданными. В структуре научного текста это находит отражение в виде зачина статей, книг, в котором их новая тема соотносится с той или иной общей проблемой, с кругом проблем в этой области. Тем самым эта новая работа вводится в рамки уже существующих проблем. Сюда же относится умение оперировать основными моментами и понятиями в истории «своей» науки, владение общим концептуальным фоном независимо от того, цитируются эти работы или нет. Именно история науки, состояние разработки той или иной проблемы не только формируют научное знание, о чем сказано было выше, но и диктуют и во многом предопределяют и характер авторского изложения, и структуру текста. Лики ученых-предшественников незримо присутствуют в каждой серьезной научной работе. В целом все это и составляет пресуппозицию письменного текста.

В самом тексте отмеченные моменты находят отражение в обзоре литературы, в системе отсылок и сносок, в комментариях. При этом в монографиях подобный обзор может занимать целую главу или быть рассеянным по отдельным главам, не говоря уже о работах по истории науки. В заключении полученные новые результаты так или иначе соотносятся с достижениями предшественников. В целом сюда же относятся и обычные указания на новизну, на защищаемые положения, идеи исследования. Эти части книг, статей изобилуют фами-

лиями ученых, географических названий, наименованиями работ, датами.

Тесно связан с последним фактором формирования научного текста и императив традиций самого текста. В целом каждый исследователь строит композицию своих работ так, как принято в этом жанре исследований, в данной отрасли знания (статья, монография, обобщающие сводки, пособия). Так, например, все общие определители по зоологии, ботанике имеют достаточно строгую структуру, при которой в определенной последовательности следует описание классов, семейств, родов, видов, подвидов, их строения, образа жизни. В географии при физико-географической характеристике отдельных районов сначала дается характеристика рельефа, геоморфологии, затем климата, далее вод, почв, ландшафта, истории географической среды.

Отмеченные выше факторы (прагматика, коммуникативная среда, традиции текста) находят свое выражение в текстах на разных языках в различных собственно языковых формах. Здесь, в языковой структуре текста, вместе со значениями слов, морфем, конструкций они и формируют целостный смысл научного текста.

Дифференцированный подход к языку научных и технических произведений с учетом макроструктуры текста особенно важен при анализе текстов, терминологии обобщающих фундаментальных работ по физике, биологии, географии, химии, в которых встречается специальная лексика разных тематических групп. В то же время эта лексика отнюдь не образует некий сплошной терминологический поток или хаос, а распределяется именно в зависимости от типов макроструктуры текста. Так, например, в общих работах по географии при описании геоморфологии рельефа используются одни термины, при описании климата или видов болот, лесов, полей — другие. Свое распределение терминов в зависимости от структуры текста выступает даже и внутри отдельных частных дисциплин. Так, в озероведении при описании озер сначала дается их физико-географическая характеристика, в которой преобладают термины геологии, гидрографии, физической географии, а затем используется своя, особая терминология в разделах по гидробиологии, ихтиофауне озер и их использованию. За общими названиями терминологических словарей биологических или физических терминов скрывается огромный мир понятий.

Каковы же составляющие компоненты структуры научного текста? Это, во-первых, его макроструктура и общая архитектоника (композиция, типы и расположение разделов, глав, параграфов, абзацев, схем, рисунков, карт), во-вторых, лексика, как общеязыковая, общенаучная, так и узко терминологическая, в-третьих, морфологические и словообразовательные средства, в-четвертых, синтаксис, в-пятых, семантика, пронизывающая все компоненты и части текста.

Таким образом, комплексный филологический анализ научного и технического текста предусматривает изучение характера его структуры, широкий историко-культурный концептуальный анализ произведений, документов, моделирование семантики и выявление всех формальных языковых средств, служащих для выражения семантики этих текстов.

В информатике глубинный детализированный подход с учетом разных видов терминов в той или иной области особенно проявляется при разработке проблемно-ориентированных автоматизированных фактографических и экспертных систем.

Сложившаяся внутренняя структура текстов должна более внимательно учитываться и в ходе работ по унификации и стандартизации терминов.

Среди многих прикладных проблем, связанных с анализом специального текста, особо выделяется задача создания полнотекстовых баз данных на ЭВМ. Любая база данных вместе со средствами ее реализации и системой управления может рассматриваться как информационная система. Для краткости назовем полнотекстовые базы данных просто базами текстов, имея в виду, что это те элементы информационных систем, в которых в качестве данных выступают любые целостные тексты на языке оригинала и полные тексты их переводов. Тогда база текстов — это автоматизированный фонд хранения, например, полных текстов статей, монографий, словарей и справочников. Существенной особенностью базы текстов является прямой ввод в нее большинства текстов (исключая памятники древней письменности) без предварительной их обработки.

Зададим себе вопрос: зачем специалист-исследователь обращается в научную библиотеку, в ее систематический каталог, в справочный отдел или к выставке новых поступлений? Чаще всего для того, чтобы: а) подобрать литературу по интересующему его вопросу; б) уточнить выходные данные книги, статьи; в) кратко ознакомиться из энциклопедий, словарей, справочников с основными проблемами своей работы; г) узнать из реферативных журналов, из изданий, экспресс-информации, каталогов издательств, обзоров о последних достижениях в той области знания, в которой он работает. Однако в конце концов любой серьезный специалист неизбежно обращается к основным фондам библиотек (книжному, журнальному, газетному и др.), т. е. к полным текстам тех самых статей, книг, о которых он узнал в ее справочных службах.

Нетрудно заметить, что базы текстов представляют собой прямой автоматизированный аналог библиотечных фондов текстов-первоисточников. При этом вряд ли приходится сомневаться в том, что в любой нашей библиотеке главное место занимают именно основные фонды и подфонды полных текстов (книг, журналов, газет, рукописей, словарей), систематизированные в соответствии с принятой классификацией.

По-видимому, вполне логично предположить, что в перспективе в базах данных библиотек основное место займут базы текстов.

Для специалистов разных областей науки, техники, управления полные тексты — это единственная возможность постичь суть того или иного понятия не из вторых рук, а непосредственно по первоисточнику. Вряд ли следует много говорить о том, что никакой реферат, никакая экспресс-информация не могут заменить текст первоисточника. Реферат не может дать подлинного представления о самой научной публикации. В нем нарушена сложившаяся в тексте первоисточника система семантических отношений термина, а тем самым — и всех понятийных связей; критерии отбора отдельных терминов в реферат не всегда бесспорны. Любые научные и технические понятия существуют только в формах того или иного естественного или искусственного языка, и проникнуть в их суть можно лишь через целостный текст первоисточника.

Развитие персональных ЭВМ и хранение полных текстов в базах текстов предоставляют любому специалисту возможность неоднократного прямого обращения к статье, книге, для разработки и поиска новых нюансов, оттенков в трактовке тех или иных понятий и решений.

Совершенно очевидны преимущества полных текстов для автоматизации филологических работ, в том числе и собственно лингвистических исследований. Прямо из текстов, введенных в ЭВМ, мы можем получить:

— полный список всех словоформ для отдельного текста, памятника, произведения;

— полный список всех словоформ для всех текстов, введенных в ЭВМ;

— элементы морфологии текста;

— частоту встречаемости каждой словоформы как в каждом из текстов, так и суммарно — во всех текстах, введенных в ЭВМ;

— адреса всех словоформ (номера глав, страниц, строк);

— типологию и статистику букв и буквосочетаний;

— контекст (контексты) разной длины для каждой словоформы как по отдельному тексту, так и по всем текстам вместе взятым;

— обратный словарь (словари);

— сам текст данного памятника (произведения) в полном виде.

Конечно, каждая новая прикладная задача требует своего подхода к специальному тексту. Однако есть некоторые моменты, которые так или иначе проявляются при решении большинства задач. К ним относятся: а) анализ макроструктуры текста; б) моделирование семантики; построение по текстам логико-понятийных систем и терминосистем: в) разработка графического представления данных; г) создание словарей разных ти-

лов; д) моделирование отдельных видов грамматических отношений.

Литература

Богданов В. В. Текст и текстовое общение. СПб., 1993.

Гальперин И. Р. Текст как объект лингвистического исследования. М., 1981.

Головин Б. Н., Кобрин Р. Ю. Лингвистические основы учения о терминах. М., 1987.

Даниленко В. П. Русская терминология. М., 1977.

Денисов П. И. Принципы моделирования языка. М., 1965.

Закономерности структурной организации научно-реферативного текста. Киев, 1981.

Лекомцев Ю. К. Введение в формальный язык лингвистики. М., 1983.

Лотман Ю. М. Структура художественного текста. М., 1970.

Лотте Д. Основы построения научно-технической терминологии. М., 1961.

Митрофанова О. Д. Язык научно-технической литературы. М., 1973.

Мурзин Л. Н., Штерн А. С. Текст и его восприятие. Свердловск, 1991.

Никитина С. Е. Семантический анализ языка науки. М., 1987.

Петров В. В. Семантика научных терминов. Новосибирск, 1982.

Реформатский А. А. Что такое термин и терминология: Вопросы терминологии. М., 1961.

Сборник рекомендуемых терминов. Вып. 85: Термодинамика. Терминология. М., 1973. 18 с.

Татаринов В. А. История отечественного терминоведения. Т. 1. М., 1994; т. 2, кн. 1. М., 1995.

Якубайтис Т. А. Части речи и типы текстов. Рига, 1981.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Электронно-вычислительная машина (ЭВМ) в современном понимании этого словосочетания впервые увидела свет в 1946 г. (ЭНИАК, США), и это явилось началом эпохи бурного развития вычислительной техники, которая продолжается и в настоящее время.

Первые вычислительные машины предназначались для выполнения объемных арифметических расчетов, т. е. они рассматривались как гигантские программируемые калькуляторы. Так, ЭНИАК проектировался для расчета баллистических таблиц. Однако уже в начале 50-х годов нашего столетия стало ясно, что использование ЭВМ не должно ограничиваться этой достаточно узкой областью, что вычислительные машины могут обрабатывать не только числовую информацию. Строго говоря, последнее высказывание не совсем точно. Правильнее же будет сказать, что практически любая информация (текст, изображение, звуковой сигнал и т. д.) может быть закодирована при помощи чисел, которые и будут обрабатываться ЭВМ. Но это замечание не слишком существенно для практической деятельности, так как при современном состоянии развития вычислительной техники и методов ее использования можно написать программу для ЭВМ, скажем, решающую задачу какой-либо обработки текста, не имея ни малейшего представления о том, каким образом этот текст будет закодирован.

Быстрому осознанию возможностей вычислительной техники (в частности, возможности обрабатывать нечисловую информацию при помощи ЭВМ) способствовало опережающее развитие математической логики. Основополагающие результаты теории алгоритмов (раздела математической логики), такие, как машина Поста, машина Тьюринга, рекурсивные функции, нормальные алгорифмы Маркова, тезис Черча — Тьюринга и т. д., были получены еще до появления первых вычислительных машин.

Таким образом, начиная с 50-х годов XX в. и до наших дней ЭВМ широко используются для решения различных задач по обработке нечисловой информации.

Тогда же, в 50-х годах нашего столетия, и был поставлен вопрос: можно ли научить машину мыслить как человек, или, другими словами, можно ли создать искусственный интеллект (ИИ)?

Вообще говоря, вопрос о создании искусственного мыслящего существа не раз возникал в истории развития человеческой мысли (от попыток алхимиков получить гомункулуса, древних и средневековых легенд до роботов Карела Чапека и Айзека Азимова), но только после появления первых вычислительных машин этот вопрос прозвучал не на уровне абстрактных философских схем и художественных произведений, а с учетом реальной, хотя и достаточно отдаленной перспективы. Впервые этот вопрос четко сформулирован и подробно рассмотрен в статье замечательного английского математика А. Тьюринга, вышедшей в 1950 г., которая так и называется — «Может ли машина мыслить?».

Ответ на вопрос, можно ли создать ИИ, получить далеко не так просто, как это может показаться на первый взгляд. Сложность этого вопроса связана с тем, что мы очень плохо представляем себе, как человек мыслит. Общей теории мышления в настоящее время не существует. Более того, мы даже не знаем, что такое мышление, не существует определения мышления, по крайней мере такого, которое удовлетворило бы одновременно философов, психологов, инженеров, лингвистов, а также других специалистов, которые участвуют в решении задач ИИ.

Предположим, что построена система, которая, как утверждают создатели, мыслит, т. е. является системой ИИ. Как проверить, «мыслит» она или нет? В чем может состоять критерий интеллектуальности? Такой критерий был предложен А. Тьюрингом в его уже упоминавшейся пионерской статье. В научной литературе этот критерий получил название «тест Тьюринга». Кратко он может быть сформулирован так: если человек будет некоторое время общаться с машиной и не сможет заметить, что он общается с машиной, а не с человеком, то можно признать, что эта машина мыслит. При этом предполагается, что разговор человека с машиной будет вестись не впрямую, а при помощи какого-либо устройства для передачи информации, например телетайпа. Это ограничение вполне естественно, так как оно исключает из критерия интеллектуальности такие параметры, как «внешний вид», «тембр голоса» и т. д. (Здесь приведено описание упрощенного теста Тьюринга. В таком виде его рассматривают большинство современных исследователей. Сам А. Тьюринг приводил этот тест в несколько ином виде, но тем не менее не возражал против такой его интерпретации.)

Несмотря на свою субъективность тест Тьюринга как критерий интеллектуальности обладает большими достоинствами. Он опирается на сравнение с человеком. Пока у нас нет общей теории мышления, признать нечто мыслящим мы можем, только сравнив это «нечто» с человеком — единственным существом, обладающим этой характеристикой априори.

Кроме того, этот тест достаточно универсален. Метод вопросов и ответов позволяет охватить практически любую область человеческой интеллектуальной деятельности. Накладывая на тест Тьюринга дополнительные ограничения, мы можем получать критерии частичной интеллектуальности, т. е. проверять решение частных задач ИИ. Например, мы можем потребовать, чтобы беседа машины и человека не выходила из рамок определенной предметной области, накладывать ограничения на используемый язык (например, на лексику или синтаксис) и т. д.

Тест Тьюринга не дает почти никаких подсказок по поводу того, какие именно частные задачи необходимо решить для того, чтобы создать машину, способную преодолеть этот тест. Поэтому круг исследуемых вопросов обычно выделяется интуитивно. И хотя различные исследователи по-разному определяют набор таких задач, тем не менее существует «устойчивая сердцевина» — проблемы, которые практически всеми специалистами включаются в число задач ИИ. К числу таковых относятся: решение задач, принятие решений, распознавание образов.

Решение задач. Этот термин объединяет многие проблемы: от доказательства теорем высшей алгебры и решения задач из учебника физики VI класса средней школы до решения бытовых проблем, например, нахождения ответа на вопрос, какой можно приготовить обед из имеющихся продуктов. Разумеется, здесь имеется в виду не непосредственное решение упомянутых выше задач, а нахождение общего механизма, который использует человек для решения подобных задач, т. е. разговор идет не о решении задач, а о моделировании процесса решения задач.

Пока исследуются вопросы, в основном связанные с моделированием решения различных математических задач. Это связано в первую очередь с тем, что математические задачи проще поддаются формализации. Кроме того, для общего решения проблемы следует разобраться с тем, каким образом человек проводит рассуждения, осуществляет логический вывод. Для моделирования решения только математических задач можно использовать формальную (математическую) логику. Однако для общего решения этой проблемы математическая логика (или, точнее, только математическая логика) вряд ли подойдет: человек пользуется логикой, весьма отличающейся от формальной логики. Это не значит, что люди не пользуются формальной логикой (или чем-то вроде этого), но они явно используют

что-то еще для логических преобразований. Вероятно, мир, где господствовала бы только формальная логика, был бы очень скучен. Но если математическая логика — это хорошо разработанная научная дисциплина, то «логика человека» остается почти не исследованной.

Проблема **принятия решений** в основном рассматривается на материале стратегических игр, таких, как шахматы, шашки, го и т. д. Работы по моделированию игр осуществляются отнюдь не только для развлечения скучающей публики. Скорее, наоборот. Коммерческие программы, играющие в шахматы и другие игры, являются «побочным эффектом» работ такого рода. Дело в том, что «правила игры» существуют не только для перемещения фигур по шахматной доске, но и для размещения сырья и оборудования по предприятиям, передвижения полков и армий по театру военных действий, перемещения денежных средств по финансовой системе и т. д.

Выработка общих методов принятия решения о выборе в каждом конкретном случае тех правил, которые приводят к наилучшим результатам (т. е. нахождение оптимальной стратегии), привела бы не только к возможности создания эффективных автоматизированных систем управления, экспертных систем, но и внесла бы определенный вклад в общую теорию мышления.

Распознавание образов. Человеку свойственно классифицировать окружающий его материальный мир. Бесчисленное множество различных объектов человек объединяет в группы. Два различных яблока, не очень похожих друг на друга, человек называет одним словом — *яблоко*. А груша, может быть, очень похожая на яблоко, тем не менее — *груша*.

Мы узнаем приятеля, вернувшегося из отпуска, хотя он загорел, отпустил бороду и надел черные очки (т. е. его внешность сильно изменилась).

Одна буква, написанная разными людьми (да и одним человеком в разное время), выглядит по-разному, но мы каким-то образом определяем, что это одна и та же буква.

Одно и то же слово, произнесенное разными людьми, звучит по-разному, оставаясь одним и тем же словом.

Какими же неуловимыми признаками пользуется человек, чтобы отождествить два разных объекта (т. е. выяснить, что эти два предмета — оба яблока, что тот Петя, которого мы видели вчера, и тот, кого мы видим сегодня, — один и тот же человек, и т. д.)?

Предполагается, что в процессе мышления человек создает некие эталонные образы объектов и групп объектов, с которыми он имеет дело, и, сравнивая образ рассматриваемого объекта с эталонными образами, относит этот объект к той или иной группе. Это действие и называется «распознавание образов». В данном контексте следует понимать слово «образ» не как некоторый психологический или нейрофизиологический термин,

а интуитивно, как некоторое представление человека об объектах, о типичных характеристиках групп объектов. В конкретных исследованиях понятие «образ» может быть уточнено.

Самое главное для решения проблемы распознавания образов — это найти ответ на два очень сложных вопроса: *как создается эталонный образ* (т. е. каким образом человек членит, классифицирует мир)? и *какие различия образа объекта с эталоном являются существенными, а какие — нет?*

Человек имеет различные каналы получения информации об окружающей его действительности — зрительный, слуховой, осязательный и т. д., поэтому у него могут формироваться образы различных типов, и проблема распознавания образов, таким образом, распадается на ряд проблем: распознавание зрительных образов, распознавание акустических образов.

Выше было сказано, что тест Тьюринга не дает почти никаких подсказок о том, какие именно задачи необходимо решить для создания ИИ. Но тем не менее одно из наиболее важных направлений проблемы ИИ он указывает: необходимо обеспечить общение человека с машиной на естественном языке. Актуальность этой задачи состоит не только в том, что любое продвижение вперед в этой области приближает нас к созданию ИИ и вносит вклад в развитие общей теории мышления, но и в том, что даже частичное, с существенными ограничениями решение этой задачи может быть использовано для многочисленных практических применений, таких, как создание естественно-языковой надстройки для информационно-поисковых диалоговых систем, автоматизированных систем управления и других систем, для эксплуатации которых сегодня требуется изучать специальные языки, пользоваться специальными словарями и т. д.

Таким образом, из широкого спектра задач ИИ важнейшей для прикладной лингвистики является проблема моделирования речевой деятельности человека, проблема анализа и синтеза речи.

Процесс общения человека и машины на естественном языке можно представить себе так: машина получает текст и, восприняв его, вырабатывает определенную реакцию. Реакция может быть различной в зависимости от назначения и возможностей соответствующей системы. Это может быть ответ на вопрос (синтезированный машиной текст), просто изменение внутреннего состояния системы (машина «принимает к сведению» информацию, содержащуюся в тексте, до поры до времени не реагируя на нее) и даже некоторое действие (робот выполняет приказ «хозяина»). Реакции последнего типа не вписываются в границы лингвистической проблематики, поэтому здесь о них речь не пойдет. Мы ограничимся проблемами вопросно-ответных систем.

При восприятии текста для того, чтобы выдать «осмыслен-

ную реакцию» на входной текст, машина должна «понять» его. Возникают вопросы: *что такое понять текст?* и *как определить, что текст понят?* Эти вопросы весьма сложны. Так же как нет определения термина «мышление», нет и определения термина «понимание». Есть только предположение, что эти понятия близки между собой.

Критерий понимания был выработан исследователями, вероятно, по аналогии с тестом Тьюринга: машина понимает текст, если реагирует на него адекватно, правильно. Субъективизм этого критерия (как и теста Тьюринга) очевиден: кто будет определять, правильно или неправильно реагирует машина на текст? Кроме того, ясно, что в этом критерии под пониманием подразумевается правильное понимание, т. е. перепутаны противопоставления «понимание/непонимание» и «правильное понимание/неправильное понимание». Однако нельзя не признать, что при создании практических систем ИИ этот критерий необходим — кому нужна, скажем, вопросно-ответная система, которая (с точки зрения пользователя этой системы) неправильно отвечает на его вопросы?

Так же как и в случае с тестом Тьюринга, этот критерий не дает подсказки, что именно необходимо сделать, чтобы текст был понят. У разных исследователей существуют различные подходы к решению этой проблемы, но в целом общее направление можно охарактеризовать так: система ИИ должна преобразовывать текст в некоторую структуру внутреннего представления (структуру представления смысла текста), на основании которой система сможет выполнить свою «интеллектуальную функцию», т. е. сформировать ответ, осуществить какое-либо действие и т. п.

Никаких общепринятых правил и установок по поводу того, как должна быть организована структура внутреннего представления, нет. Есть только одно достаточно нечеткое требование, чтобы эта структура адекватно отображала смысл введенного текста и была бы пригодна для осуществления различных «интеллектуальных» действий (например, для логического вывода). Разным смыслам должны соответствовать разные структуры.

Кроме того, некоторые исследователи предъявляют требование, чтобы процесс понимания текста системой ИИ происходил «так же, как и у человека». Они считают, что в противном случае успеха добиться невозможно. Определенная логика в этом есть. Человек — единственное сегодня существо, которое может гордиться тем, что умеет понимать текст; поэтому тот процесс, который происходит в голове у человека, сегодня является единственным, приводящим к положительным результатам в этой области. Однако требование, чтобы в системе ИИ все происходило, «как у человека», вряд ли реально. Каждый исследователь имеет свое представление о том, как именно у человека происходит процесс понимания, но едва ли кто-

нибудь из них сможет убедительно обосновать свои представления. Голова человека является для исследователей «черным ящиком», а достижения нейролингвистики не столь велики, чтобы мы могли сегодня достаточно четко представлять, что происходит в голове в процессе понимания. Поэтому большинство специалистов хотя и пытаются в своих работах реализовать свои представления о том, «как это происходит у человека», тем не менее в публикациях осторожно говорят о функциональном моделировании речевой деятельности.

Итак, для того чтобы текст был понят системой ИИ, он должен быть переведен в структуру внутреннего представления, которая отражает смысл текста. Для осуществления этого перевода система ИИ должна использовать некоторую информацию, которую специалисты по ИИ обычно называют *знаниями*. Общепринято разделение *знаний* на *знания о языке* и *знания о мире*. Последние называют еще *энциклопедическими знаниями*.

Провести точную границу между этими двумя типами знаний достаточно сложно (может быть, даже невозможно). Каждый исследователь решает этот вопрос по-своему. Обычно к *знаниям о языке* относят морфологическую, синтаксическую и семантическую информацию, т. е. те сведения, которые позволяют определить структуру текста и основные связи между словами. Вся остальная информация, необходимая для понимания текста, относится к *знаниям о мире*.

Создатели первых экспериментальных систем, понимающих текст на естественном языке, пытались обойтись только *знаниями о языке*, но вскоре выяснилось, что этого недостаточно. Рассмотрим для примера два предложения: *Городские власти запретили шествие демонстрантов, так как они опасались насилия* и *Городские власти запретили шествие демонстрантов, так как они призывали к революции*. С точки зрения языка оба предложения построены приблизительно одинаково, и *знания о языке* не помогут нам разрешить проблему референции местоимений, т. е. выяснить в каждом из этих предложений, кто такие «они». Для этого необходимы знания об окружающем нас мире, об отношении городских властей к демонстрантам, революции, насилию и т. д., т. е. те знания, которые никак нельзя назвать *знаниями о языке* (пример американского ученого Т. Винограда).

У различных людей (а также, может быть, и у различных систем ИИ) *знания о мире* могут достаточно сильно различаться. А как мы выяснили, эти знания существенным образом влияют на понимание текста. Поэтому разные люди могут по-разному понимать один и тот же текст. Это еще одно подтверждение субъективности приведенного выше критерия.

Проиллюстрировать различное понимание одного текста разными людьми можно достаточно простым примером из некогда популярной песни, в которой есть такие слова:

...— я у аптеки...
— А я в кино искала Вас...
— Так, значит, завтра
На том же месте, в тот же час!

Герои этой песни никак не могут встретиться потому, что поразному понимают текст. При этом даже нельзя сказать, кто из них понимает правильно, а кто — нет. Каждый из них посвоему прав. Просто они обладают различными *знаниями о мире*. (В данном случае — знаниями о том, что можно считать «тем же местом».)

Для того чтобы понять даже очень простой текст, нужно использовать большое количество *знаний о мире*. И если *знания о языке* можно более или менее полно описать (а именно этим занимаются лингвисты), то *знания о мире* потенциально бесконечны (невозможно знать все обо всем). Кроме того, познания человечества *о мире* с каждым днем расширяются. Поэтому при создании систем понимания появляются принципиальные проблемы, связанные со *знаниями о мире*. Всех знаний задать невозможно, а какие именно понадобятся в процессе работы, предугадать тоже очень сложно. Основная часть создающихся систем ориентируется на очень узкую предметную область, т. е. такие системы могут понимать тексты ограниченной тематики. Но даже при таких ограничениях результаты очень далеки от идеальных.

Другой подход связан с созданием обучающихся систем. Этот подход представляется более перспективным. При этом подходе система понимания должна уметь извлекать из текста новую информацию и присоединять ее к имеющимся у системы *знаниям о мире*. При нехватке информации для понимания текста система может задавать вопросы человеку по поводу неясных мест и таким образом постепенно накапливать *багаж знаний*. Системы, обладающие одинаковым набором знаний в начале эксплуатации, могут прийти к различным результатам в зависимости от того, как и чему их учили.

Но вне зависимости от того, каким образом пополняются или не пополняются *знания о мире*, главная задача при понимании текста — это создать структуру внутреннего представления текста, т. е. структуру, отражающую смысл текста. Обычно эта структура строится при помощи процедур, аналогичных тем, которые используются для других задач автоматической обработки текста (машинный перевод, автоматическое реферирование и т. п.), а именно — процедур морфологического, синтаксического и семантического анализа.

Построенная структура используется для выработки реакции системы (формулирования ответа, выбора действия и т. п.). Следует отметить, что изучение этого этапа — задача не только для лингвиста, но и для представителей многих других дисциплин, например, психологии (определение модели поведения), ло-

гики (*осуществление логического вывода*) и т. д. Практически все направления исследований в области ИИ затрагивают интересы различных специальностей и поэтому должны вестись комплексно, с привлечением специалистов разных дисциплин.

Все описанные выше проблемы в настоящее время не имеют общих решений. Правда, существуют весьма многочисленные попытки решения (с бо́льшим или меньшим успехом) частных очень узких задач — если осуществляется диалог человека с ЭВМ на естественном языке, то обычно только в рамках весьма ограниченной предметной области, если моделируется принятие решений, то только в пределах конкретной задачи (например, выбор хода в шахматной партии) и т. д.

Вероятно, можно сказать, не погрешив против истины, что общая теория мышления сегодня только начинает создаваться.

Литература

Моделирование языковой деятельности в интеллектуальных системах / Под ред. А. Е. Кибрика и А. С. Нариньяни. М., 1987.

Попов Э. В. Общение с ЭВМ на естественном языке. М., 1982.

Фитиалов С. Я. Понимание естественного языка в системах искусственного интеллекта // Лингвистические проблемы функционального моделирования речевой деятельности. Вып. V. Л., 1982. С. 3—9.

Хант Э. Искусственный интеллект. М., 1978.

МОДЕЛИРОВАНИЕ ЯЗЫКА

Изложению конкретных проблем моделирования языка целесообразно предпослать введение, в котором рассматривается ряд общих вопросов. Прежде всего следует дифференцировать три различных, хотя и тесно связанных друг с другом феномена.

1. Речевая деятельность человека (или, иначе, речевое поведение), которую можно свести к трем основным типам: а) обычное общение (коммуникация) людей в процессе человеческой деятельности и жизни; б) перевод с одного языка на другой; в) обучение языку или любым другим знаниям из различных областей науки, культуры, искусства и т. д.

2. Во всех типах речевого поведения общающиеся между собой люди (коммуниканты) имеют дело с пониманием и продуцированием устных и письменных речевых произведений (далее для их обозначения будет также использоваться условный синоним тексты).

3. Тексты понимаются и продуцируются человеком на основе знания языкового механизма данного языка, т. е. знания лексического состава языка, правил построения и изменения слов, правил построения предложений, целого текста и т. д.

На основе приведенного деления можно говорить о трех типах моделей языка.

Модели, воспроизводящие тот или иной тип речевой деятельности. Это модели понимания (анализа) текста, модели порождения речевых произведений (модели синтеза), модели автоматического перевода, обучающие модели языка, интеллектуальные системы разных типов, использующие диалог на естественном языке и т. д.

Второй тип моделей — речевые произведения. Специфика роли речевых произведений в аспекте проблем моделирования заключается в следующем: речевые произведения (тексты) являются обязательными компонентами любых типов речевой дея-

тельности, а стало быть, и обязательными компонентами любых ее моделей.

В то же время, как мы попытаемся показать ниже, речевые произведения сами являются некоторой моделью действительности, моделью мира, и этот факт, имеющий большое значение для создания моделей первого и третьего (см. ниже) типа, должен быть учитываем конструкторами любых языковых моделей.

Третьим типом моделей являются модели собственно языкового механизма, которые воспроизводят языковые закономерности понимания и построения речевых произведений.

Некоторые положения из сказанного необходимо развить далее.

1. Естественный язык как модель мира

Тезис о том, что естественный язык моделирует окружающий нас мир, постулируется многими современными лингвистами. Однако представляется, что этот тезис справедлив лишь в том случае, когда речь идет о языке, представленном в виде речевых произведений — текстов. Языковая система (языковой механизм) также участвует в моделировании (отражении) мира, но это участие опосредованное.

Положение о том, что речевые произведения моделируют мир, можно подтвердить анализом общих свойств моделей и выяснением того, отвечают ли речевые произведения этим свойствам.

Напомним, что в общем случае модель понимается как некая конструкция, в которой некоторым способом воспроизведена та или иная система действительности.

Перечислим ряд общих свойств моделей и ситуации моделирования, существенно влияющей на эти свойства, и сопоставим их со свойствами языка как модели мира и ситуацией моделирования, характерной для этого случая.

1) Дуализм ситуации моделирования. Любая ситуация моделирования очевидным образом предполагает наличие моделируемого объекта (воспроизводимой конструкции — Поспелов, 1979) и модели (воспроизводящей конструкции). В языке моделируемым является окружающая человека действительность, а моделью — тексты, ее описывающие.

Специфика языка как модели мира заключается в этом случае в расширенном понимании действительности, в понятие которой включается не только окружающий нас физический мир, но и интеллектуальная, духовная, социальная жизнь человека, а также все ценности, созданные человечеством, в том числе и сами речевые произведения. Специфичность языка как модели мира состоит также в почти неограниченной способности текстов моделировать нереальный мир (согласно различной терминологии, возможный, мыслимый, вероятностный, фиктивный, проблемный). Как писал известный французский лингвист Ш. Балли, «действительность может быть не только объектив-

ной, но и мыслимой, воображаемой» (Балли, 1955, с. 88). Эта способность реализуется в сказках, легендах, мифах, фантастике и даже в реалистичных произведениях художественной литературы, которая тем не менее отражает действительный мир опосредованно, при помощи типизированных образов не существующих в реальной жизни персонажей и выдуманного, как правило, сюжета.

2) Системность моделируемого. По определению, модель воспроизводит некоторую систему, состоящую из элементов или частей, взаимосвязанных друг с другом и представляющих одно целое и/или функционирующих как одно целое.

Речевые произведения — тексты также воспроизводят в речевой форме системы действительности. Специфика заключается в возможности описывать не только очень большие системы, например строение космоса, но и очень малые системы, например, микромиры повседневной жизни человека, описываемые краткими диалогами.

3) Системность модели. Любая модель представляет собой систему элементов, связанных между собой определенными отношениями.

Каждый текст также является системой элементов, связанных друг с другом смысловыми, структурными и линейными отношениями.

4) Аспект воспроизводимости моделируемого в модели. Аспект воспроизводимости зависит от задач и целей моделирования. В зависимости от того, какой аспект моделируемого воспроизводится в моделях, последние делятся на структурные, функциональные и смешанные модели. В структурных моделях воспроизводится структура моделируемого объекта, т. е. его существенные элементы в системе их отношений друг с другом. Функциональные модели воспроизводят поведение объекта в природе и обществе. В смешанных моделях объединены те и другие аспекты.

Речевые произведения относятся к числу структурных моделей особого типа. Даже в случае, если речевое произведение описывает систему функционирования какого-либо кусочка действительности, оно тем не менее не является функциональной моделью, являясь структурной моделью функционирования.

5) Тип модели. Модели могут быть интеллектуальными и физическими. Речевые произведения — интеллектуальные структурные модели.

6) Дуалистический характер модели. Любая модель представляет собой некое единство формы и содержания и, тем самым, имеет дуалистический характер. Следует заметить, что форма и содержание модели отличны от формы и содержания моделируемого объекта, даже в случае сильной имитации. Например, форма модели корабля отличается от натурного объекта размером, степенью детализации составных частей, мате-

риалом и т. п. Формой интеллектуальных моделей являются описания, сделанные при помощи естественного и искусственных языков, а их содержанием — система элементов и отношений между ними.

Что касается текстов как моделей мира, то их формой следует считать временную (для устной речи) или линейную (для письменной речи) последовательность текстовых элементов, а содержанием — грамматическую и смысловую структуру текста. Дуалистический характер речевых произведений, очевидно, объясняется знаковой природой языка.

7) Обобщение и упрощение моделируемого в модели. *Модель — всегда каким-то образом обобщенное и уже тем самым* упрощенное воспроизведение структурных, функциональных или тех и других свойств моделируемого явления. «Под моделью понимается обобщенное описание объекта или класса объектов, при котором некоторые параметры остаются неопределенными» (Гузман, 1975, с. 163).

При помощи естественного языка мир моделируется с различной степенью точности в зависимости от целей коммуникации, однако всегда достаточно упрощенно и обобщенно. Общеизвестны наблюдения лингвистов и логиков относительно приблизительности и неопределенности значения языковых выражений. Например, неопределенно значение множественного числа: о денотативном статусе субъекта в предложении *Малыши играли в саду* точно можно сказать только, что малышей было не меньше двух, а в случае употребления кванторных слов *некоторые, несколько* — что каких-то предметов или лиц было больше двух. Выражение *Мальчики дрались друг с другом* может моделировать несколько ситуаций реальной действительности, среди которых может быть и такая, когда двое мальчиков дрались с третьим, несколько сделали «кучу малу», а один не принимал активного участия в драке, но вертелся среди дерущихся. Для каких-то целей коммуникации такой точности сообщения оказывается достаточно. Например, сообщение: *Тетя Галя, мальчишки дерутся друг с другом, и ваш Максим там!*, — не требуя уточнений, вызывает определенную реакцию матери, т. е. выполняет свое коммуникативное назначение. Для задач же, требующих уточнения, например, при составлении протоколов нарушения порядка, применяются дополнительные речевые процедуры выяснения истины.

Одной из кардинальных причин неточности и неопределенности языковых выражений служит свойство асимметрии языкового знака, присущее языковому механизму на любом уровне его структуры и закономерно проявляющееся в текстах, образуемых при помощи этого языкового свойства. Асимметричные элементы словаря — синонимия и омонимия характерны для любого естественного языка. Однозначное соответствие отсутствует также в грамматических категориях. Например, в грам-

матических системах естественных языков имеется довольно ограниченное количество специальных средств выражения временных отношений, в то время как логическое их исчисление превышает полторы сотни (Поспелов, 1979). Примером обобщающей функции языка может служить сам процесс номинации. Как известно, носители языка называют словом *стол* весьма разнообразные конструкции, объединяемые, правда, общностью функций, обобщая на основе индивидуального и коллективного опыта функциональные и структурные особенности любых столов.

Специфика языковых выражений такова, что обобщенность, неточность и неопределенность в общем случае не уничтожает возможности успешной коммуникации, тем более что достижение успешного результата общения может быть обеспечено дополнительным речевым общением.

8) Отличие модели от моделируемого объекта. Этот тезис логически вытекает из перечисленных выше свойств. Даже модель, не созданная «руками» человека, например, маленький водоворот с параметрами движения воды, аналогичными большому водовороту, будет отличаться от второго размером, силой воздействия на окружающую среду и т. д.

Речевые произведения отличаются от описываемой ими действительности.

9) Качество модели. Модели могут воспроизводить моделируемое с различной степенью адекватности и полноты. Качество модели помимо объективных факторов (например, свойство обобщения и упрощения моделируемого) определяют также и субъективные факторы, обусловленные уровнем знаний и способностями создателей модели как в области моделирования, так и в области знания окружающего мира. Очевидно, однако, что наличие плохих моделей не может поставить под сомнение саму идею моделирования.

Тексты также могут отражать действительность с различной степенью адекватности и полноты или в некоторых случаях не моделировать ее совсем, например, при преднамеренной дезинформации. Однако эти случаи не могут поставить под сомнение сам тезис о моделировании мира текстом.

10) Активность создателей модели. Модель — это всегда создание активной интеллектуально-физической деятельности человека или коллектива людей. При моделировании действительности автором текста активно выбирается, актуализируется в речи тот ее аспект, который с точки зрения автора релевантен для целей сообщения. Выбор целей сообщения, в свою очередь, также активная, целенаправленная операция.

Специфичность речевых произведений как моделей мира заключается в этом случае в том, что языковой механизм, при помощи которого тексты порождаются и воспринимаются, — результат творчества социального коллектива, передается от по-

коления к поколению путем обучения в раннем детстве и совершенствуется в течение всей жизни человека, в результате чего создается автоматичность владения им. Таким образом, активно и сознательно моделируя действительность при помощи текстов, человек их порождает и воспринимает, почти автоматически пользуясь языковым механизмом.

11) Целеполагание в ситуации моделирования. Модель как инструмент познания действительности создается с определенными, как правило, социально значимыми целями — использовать знания о моделируемом объекте или явлении действительности, полученные при помощи модели, в различных областях человеческой деятельности.

Социально значимыми целями моделирования действительности при помощи речевых произведений являются коммуникации между людьми для осуществления человеческой деятельности в различных областях действительности. Нарушения принципа целеполагания, возможные при речевом общении, являются исключением из правил и могут быть отнесены к случаям нарушения условий успешности речевого акта.

12) Спецификация как необходимое свойство модели. Видимо, теоретически и, безусловно, практически невозможно создать модели действительности во всем ее многообразии. Моделируются обычно «кусочки», части действительности.

Речевые произведения воспроизводят также отдельные «куски» действительности.

Приведенное выше сопоставление имело целью показать, что речевые произведения — тексты моделируют действительный мир. Это положение имеет большое значение для понимания трудностей, связанных с проблемами моделирования различных видов речевой деятельности, т. е. с проблемами создания кибернетических ее моделей.[1] В частности, это положение помогает теоретически обосновать практическую невозможность создания универсальной модели, воспроизводящей все виды речевого поведения человека и анализирующей/синтезирующей любые тексты, описывающие все миры действительности как реальные, так и нереальные.

Таким образом, сложившаяся в последнее время практика моделирования систем речевого поведения, ориентированных на узкую предметную область, получает теоретическое обоснование в идеях о моделировании мира при помощи речевых произведений. Однако следует подчеркнуть, что даже при ориентации кибернетической системы на узкую предметную область эффективность и сама возможность автоматической обработки достаточно большого количества текстов сущест-

[1] Кибернетической моделью называется «некоторое (приближенное или упрощенное) представление действия конкретной исследуемой системы» (Уилсон, Уилсон, 1968, с. 19).

венно зависит от полноты, универсальности модели языкового механизма. Поэтому возникает необходимость построения общей модели языкового механизма, являющейся необходимым компонентом кибернетической модели речевого поведения.

II. Моделирование языка (языкового механизма)

Если бы в основе различных видов речевого поведения и его непременного компонента — текста не было некоей общности, то коммуникация между людьми вряд ли была бы возможной. Эта общность образуется единым для каждого конкретного языка языковым механизмом, грубо говоря — грамматикой данного языка и его словарем. Языковой механизм во всех его аспектах представляет собой систему, определенным образом устроенную и четко функционирующую, иначе он не смог бы служить людям инструментом для осуществления речевой деятельности в процессе коммуникации — люди попросту перестали бы понимать друг друга. Феномен конструирования текстов как необходимых компонентов речевой деятельности обусловливает специфику моделирования — невозможно построить модель речевого поведения, не имея какой-либо гипотезы о структуре языка, хотя бы самой элементарной. Даже в первых системах автоматического пословного перевода, где предусмотрен лишь словарь и все неточности исправляются постредактором, заданы грамматические сведения о нахождении основы текстового слова, набор грамматических значений аффиксов и флексий или парадигма словоформ класса лексем.

«Сильные» же модели речевого поведения должны предусмотреть в своем составе «сильную» модель языкового механизма, в которой должны быть воспроизведены по возможности все аспекты языковой системы.

Обычные лингвистические описания структуры языка могут быть классифицированы как интеллектуальные структурные модели. Для ряда социально значимых целей эти модели оказываются высокоэффективными. Они используются для обучения родному и неродному языку, для целей перевода, редактирования и т. п. Очевидно, что они рассчитаны на интеллект человека, способный широко использовать так называемую интуицию, т. е. ассоциации, прошлый опыт, догадки и т. п.

Следует отметить, что в плане создания таких моделей лингвистика накопила такой материал, которому может позавидовать любая другая наука и который, несомненно, должен в той или иной мере быть использован при разработке кибернетических моделей речевого поведения. Большое же количество концепций, теорий и точек зрения, которое служит иногда в качестве упрека со стороны представителей других наук, обусловлено ненаблюдаемостью языкового механизма — сведения о его функционировании и структуре исследователи получают из речевых произведений, которые, помимо всего прочего, сами

являются, как выше было показано, моделями первого порядка — моделями мира.

Модель языка, предназначенная для кибернетической системы, *воспроизводящей тот или иной вид речевого поведения* человека, очевидным образом должна быть рассчитана на «интеллект» ЭВМ, и это прежде всего диктует определенные требования к модели, отличные от требований к «человеческой» модели.

Далее речь пойдет о задачах и трудностях моделирования языка как составной и необходимой части кибернетических моделей речевого поведения. Здесь многое является общим с обычными лингвистическими исследованиями, основное отличие определяется необходимостью учета требований точности и полноты, необходимых для кибернетических моделей.

1) Активно и сознательно моделируя действительность при помощи речевых произведений, человек порождает и воспринимает их автоматически, т. е. он автоматически владеет языковым механизмом. Язык в этом случае является типичным примером кибернетического «черного ящика». Он реально существует, эффективно функционирует, однако не дан в непосредственном восприятии не только его исследователям, но и самим участникам коммуникации. Последним известно лишь то, что находится на входе и выходе этого «черного ящика». Так, в общем случае автору текста известен тот смысл, который он намерен передать реципиенту (адресату). В ситуации восприятия, наоборот, реципиенту дан текст, смысл которого он должен воспринять и понять. Следует учесть и третью позицию — позицию наблюдателя, в которой обычно оказывается исследователь. Она сходна с позицией реципиента, так как наблюдателю также дан текст, однако отличается от нее тем, что текст специально не предназначен для исследователя, и исследовательские цели, как правило, не совпадают с целями реципиента. Как устроен и как функционирует язык, исследователю также не дано наблюдать непосредственно, и свои гипотезы о структуре и функционировании языка он выдвигает лишь на основе изучения речевой продукции.

Свойство ненаблюдаемости структуры языка обусловливает основную трудность моделирования языкового механизма.

2) Идеи теоретического языкознания (Л. В. Щерба), а также имеющийся в прикладной лингвистике опыт построения кибернетических моделей речевого поведения показывают, что механизмы порождения и восприятия речи различаются исходными данными, целями, процедурами. Это различие хорошо прослеживается в моделях анализа (восприятия, распознавания, понимания) текста и моделях синтеза (порождения). В соответствии с этой дифференциацией должна строиться и интерпретация лингвистических явлений и, в конечном счете, лингвистическая теория.

В качестве примера приведем широко используемое в лингвистике понятие пресуппозиции. Оно получает другую интерпретацию в ситуации восприятия текста. Фразы *Ваня заболел гриппом, Маша уже приехала*, автор конструирует, исходя из знания о том, что Маша и Ваня существуют (так называемая экзистенциальная пресуппозиция), что, по всей вероятности, Ваня до этого момента не болел гриппом (вероятностная контрастирующая пресуппозиция), что Маша куда-то уезжала (абсолютная контрастирующая пресуппозиция). Таким образом, эта интерпретация вызвана и обусловлена позицией автора текста. Реципиент этих дополнительных сведений не имеет, а «вычисляет» на основе знания закономерностей выведения этих сведений — он делает содержательные выводы из текста. Умение делать содержательные выводы поэтому должно быть предусмотрено в любой кибернетической системе понимания текста.

В позиции реципиента-наблюдателя, на которого данный текст не рассчитан, значимость учета содержательных выводов возрастает еще более. Очевидно, что учет всех таких факторов требует более углубленных теоретических лингвистических исследований, связанных с речевым актом. В то же время повышается роль семантических исследований, в результате которых большое значение получают слова и выражения, до недавнего времени считавшиеся «второсортным» объектом исследования, как-то частицы, союзы и даже наречия. Например, на основе учета семантики слова *снова* в газетной фразе *Премьер снова посетил нашу страну* система (так же, как и человек) может сделать вполне достоверный вывод о том, что премьер посещал нашу страну, по крайней мере, еще раз.

3) Важным является установление конечной цели моделирования. Потеря «бдительности» в этом случае может привести к подмене моделирования естественного процесса коммуникации моделированием компетенции лингвиста-исследователя.

В то же время следует отметить, что моделирование деятельности лингвиста стало самостоятельной задачей компьютерной лингвистики. В качестве примера можно привести составление автоматических словарей. Такие системы являются кибернетическими исследовательскими моделями.

4) Следующая проблема моделирования языкового механизма связана с тем, что его конечной целью является получение текста, который, как мы выше пытались показать, является моделью «первого» порядка, воспроизводящей какой-либо кусочек действительности. В связи с этим проблемы задания и извлечения смысла, соотнесения содержания и формы, соотнесения содержания с действительностью, т. е. проблема референции, встают во главу угла. Появляется необходимость явно описать связь семантики и референции. Такую связь лингвисты установили давно, предложив и тщательно (а также разносторонне) разработав гипотезу о роли словаря для различных си-

туаций речевой деятельности. Вообще в естественных (не исследовательских) ситуациях речевой деятельности невозможно представить себе какой-либо язык вне словаря. Ребенок, овладевающий родным языком, человек, изучающий иностранный язык, овладевают им не абстрактно, а пользуясь конкретными словами, имеющими определенный смысл. Правильное, адекватное представление о языковом механизме создается у человека путем накопления знаний об употреблении слов. Однако представление о том, что дискретными знаками, имеющими форму и значение и образующими систему значимостей, являются слова, достаточно условно, так как такими же свойствами обладают и словосочетания. Общеизвестны в этом отношении трудности лексикографических описаний: несмотря на то, что каждый словарь строится как словарь слов, в каждой словарной статье представлен набор словосочетаний и выражений, зачастую выполняющих функции смысловых единиц. Таким образом, представление о смысловых единицах, описываемое при помощи словарей слов, — не более чем гипотеза об инвентаре знаков языка, хотя и высоковероятная.

Лексикографические трудности указанного типа весьма актуальны при создании автоматических словарей.

5) Создание правил вычисления не выраженных вербально смыслов, правил разрешения омонимии и определения синонимии слов и выражений — все это требует систематического, структурного описания семантики. Определенную часть описаний семантического представления предложения и целого текста можно причислить к модельным описаниям семантики, которые можно использовать при построении моделей речевой деятельности. Объектом моделирования стали также правила вычисления референциального значения. Особенно это актуально для экспертных систем искусственного интеллекта, использующих команды на естественном языке для исполнения интеллектуальных действий со стороны ЭВМ.

6) Создание кибернетических моделей речевой деятельности имело своим положительным результатом появление возможности проверять эффективность модели, саму возможность ее действия в отчуждении от языковой интуиции создателя модели. Немашинные модели также проверяются на практике, но эта проверка существенно зависит от языковой интуиции пользователя.

При решении автоматических задач, связанных с моделированием языка, выявилось, что имеющиеся лингвистические описания при всех их высоких качествах не могут быть использованы в не адаптированном для автомата виде. В частности, они не отвечают требованию точности. В качестве примера можно привести понятие синтаксической зависимости элементов предложения, известное в лингвистике давно (А. М. Пешковский, Д. Н. Ушаков, А. Н. Гвоздев, Л. Теньер и др.) и дав-

но применяющееся в школьном и вузовском обучении родному и неродному языку. Однако предлагающиеся схемы разбора предложения представляют собой лишь примеры, образцы для построения иерархии связей, и, как правило, без длительного обучения или дополнительных консультаций освоить их для всех типов предложений не представляется возможным. Такое положение недопустимо для постановки задачи на ЭВМ.

К описанию языка для автоматических задач стали применять точные, формальные методы, положившие начало созданию математических моделей языка. «Математическая модель языка (или языков) есть некоторая математическая конструкция, свойства которой в определенном отношении близки свойствам реального языка» (Цейтин, 1959, с. 45).

В зависимости от того, какая математическая конструкция используется, создаются различные математические модели языкового механизма (формальные грамматики). Так, для точного описания структуры зависимостей предложения была создана математическая модель, для формулирования которой был применен аппарат математической теории графов, алгебры, математической логики.

Математическое моделирование послужило стимулом для развития новых представлений о языковом механизме. Ярким примером этого может служить открытие, исследование и использование свойства проективности предложений естественного текста, которое, по мнению одного из первых исследователей этого свойства, французского ученого Ива Лесерфа, объясняет «конфликт» между линейностью элементов предложения и структурностью их синтаксических связей.

Математические модели языка стали использоваться в кибернетических моделях речевого поведения. Вначале, как известно, это были системы автоматического перевода. Сразу же выявились недостатки математических моделей языка. Жесткие математические конструкции позволяли обработку далеко не всех, достаточно характерных явлений естественного языка — неполнота модели оказалась ее неотъемлемым качеством, которое далеко не всегда оправдывается общим свойством всех моделей обобщать и упрощать моделируемый объект. Не обеспечивается свойство адекватности модели, т. е. правильности воспроизведения в ней свойств моделируемого объекта. Например, исследования показали, что большое количество предложений русского языка не отвечают свойству проективности, хотя являются правильными русскими предложениями. В связи с этим начали создавать математические модели с менее строгими правилами проективности, при помощи которых можно получить правильный автоматический анализ для предложений с разной степенью непроективности. Таким способом была повышена степень адекватности модели зависимостей. Развитие

автоматического моделирования идет по линии взаимной коррекции лингвистических и математических моделей.

Роль последних в развитии прикладной и структурной лингвистики трудно переоценить. Достаточно сказать, что благодаря им вообще стало возможно автоматическое моделирование речевого поведения человека.

7) Одна из объективных трудностей лингвистической науки в задаче моделирования связана с отсутствием достаточно полных и адекватных лингвистических моделей, воспроизводящих не статичное состояние языкового механизма, а динамические процессы речепроизводства и речевосприятия. Актуальность таких моделей объясняется тем, что в естественных условиях коммуникации все подсистемы языка действуют не изолированно и не в жестко фиксированной последовательности, как это представлено в большинстве имеющихся автоматических систем, а каким-то другим способом, например, при помощи свободных взаимных обращений одной подсистемы к другой.

Так как динамические процессы функционирования языкового механизма также являются ненаблюдаемым явлением, то очевидны трудности создания его модели. В частности, в понятие функционирования входит учет подсистемы употреблений языковых форм и конструкций. Подсистема употреблений для автоматических задач изучена недостаточно (в некомпьютерной лингвистике проблема употреблений решается при помощи теории поля).

8) Особую задачу прикладной лингвистики представляют кибернетические системы, моделирующие устное восприятие и порождение речи. Создание таких моделей невозможно без разработки структурных моделей фонологии, а также динамических моделей производства и восприятия/распознавания устной речи.

9) В кибернетических системах обычно моделируется языковой механизм в синхронном аспекте. Однако понятие синхронного среза языка, как известно, достаточно условно, потому что язык непрерывно изменяется, обслуживая многообразные сферы деятельности, контактируя с другими языками, существуя в территориальной разобщенности частей социума. Действительно, изменяется грамматика — в качестве примера можно привести замену флексий множественного числа типа *профессоры — профессора*, произошедшую в XX в. Наиболее подвижным является словарь: создаются новые слова и словосочетания, развивается полисемия и омонимия и т. д.

Для модели с высокой степенью адекватности желательно предусмотреть возможность обработки инноваций. В ином случае эта проблема попадает в число проблем, ограничивающих «силу» модели.

Резюмируя сказанное, следует еще раз подчеркнуть, что кроме создания *математических моделей языка*, являющихся компонентом *математического обеспечения*, для кибернетических моделей речевого поведения требуется разработка *структурных лингвистических моделей*, воспроизводящих процессы *функционирования* языкового механизма, а также *структурных моделей*, воспроизводящих структуру языкового механизма. В отличие от обычных лингвистических описаний эти модели должны отвечать требованиям точности, адекватности, простоты, компактности, непротиворечивости и полноты. Свойства адекватности и полноты могут быть при этом ограничены, и ограничение должно быть специально оговорено.

Литература

Балли Ш. Общая лингвистика и вопросы французского языка / Пер. с фр. М., 1955. 416 с.

Гузман А. Интегральные роботы / Пер. с англ. Вып. 2. М., 1975. 528 с.

Поспелов Д. А. Профессионально и проблемно ориентированные интеллектуальные системы // Вопр. философии. 1979. № 2. С. 57—59.

Уилсон А., Уилсон М. Информацит, вычислительные машины и проектирование систем. М., 1968. 230 с.

Цейтин Г. С. К вопросу о построении математических моделей языка // Тез. совещ. по математ. лингвистике 15—21/IV 1959 г. Л., 1959. С. 45.

ФОНЕТИЧЕСКИЕ АСПЕКТЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

Звуковая форма существования языка является основной и исходной для любого естественного языка. Это определяет особую роль фонетических исследований в прикладной лингвистике. Можно говорить о следующих основных направлениях в изучении звуковой формы языковых единиц.

1. Анализ акустических характеристик речевого сигнала — для создания систем передачи речи по каналам связи, для разработки процедур автоматического распознавания и синтеза речи, для улучшения методики восстановления речи и слуха у больных с соответствующими нарушениями.

2. Анализ артикуляторных характеристик процессов речепроизводства — для выявления механизмов управления произносительными органами со стороны соответствующих центров коры головного мозга. Это имеет большое значение для понимания процессов, характеризующих высшую нервную деятельность человека. Такое понимание имеет и практическое значение при лечении больных, страдающих различными формами нарушения речи. Многие исследователи считают сведения о механизмах речеобразования очень важными для построения систем автоматического распознавания и синтеза речи.

3. Исследование процессов восприятия речевых сообщений слуховой системой человека важно как для оценки значимости акустических свойств этих сообщений, так и для построения модели слуховой обработки речевого сигнала. Легко понять, что эти сведения чрезвычайно важны и для специалистов в области записи, передачи речевых сообщений, и для специалистов в области автоматического распознавания и синтеза речи.

Необходимо обратить особое внимание на то, что все перечисленные направления в изучении свойств речевого сигнала имеют огромное значение при обучении родному или неродному

языку. Большое место в этой проблеме занимают выработка орфоэпических навыков, устранение диалектных и просторечных явлений, а также постановка так называемого сценического произношения. Рассмотрим более подробно, какие именно проблемы встают перед прикладной фонетикой в наиболее типичных случаях.

РЕЧЕВОЙ СИГНАЛ В ТРАКТАХ СВЯЗИ И ВЕЩАНИЯ

Специалисты в области записи и передачи акустических сигналов различают первичные и вторичные акустические сигналы. Первичные — это собственно речевые сигналы (наряду с пением, звуками музыкальных инструментов и т. д.), а вторичными считают сигналы, воспроизводимые электроакустическими устройствами, т. е. «первичные сигналы, прошедшие по электроакустическим трактам связи и вещания и соответственно видоизмененные по своим параметрам» (Сапожков, 1978. С. 40). Изменения затрагивают такие свойства сигналов, как динамический диапазон, частотный диапазон, спектральные и временные характеристики. Основная задача создателей записывающих и передающих систем заключается в том, чтобы вторичный сигнал создавал у слушающего такое же ощущение, какое создает и первичный сигнал. Конечно, это достигается далеко не всегда, поэтому для оценки результатов работы, качества вторичного сигнала применяют разнообразные объективные и субъективные методы.

Объективно можно сравнивать частотные, динамические, временные характеристики создаваемой аппаратуры с соответствующими характеристиками речевых сигналов и по степени их совпадения судить о качестве вторичного сигнала. Например, если мы знаем, что динамический диапазон человеческой речи — 25—35 дБ, а динамический диапазон микрофона, с помощью которого ведется запись речевого сообщения, составляет всего 20 дБ, то легко ожидать, что вторичный сигнал будет гораздо беднее, чем естественный речевой сигнал.

Субъективными можно называть такие методы оценки вторичного сигнала, при которых используются сведения об их восприятии человеком. Конечно, для того чтобы получаемые сведения не зависели от индивидуальных свойств испытуемых или от условий эксперимента, нужно привлекать достаточно большой материал, как это обычно и делается в таких случаях, так что субъективными эти данные являются лишь потому, что в опытах участвуют люди, результаты же получаются вполне объективными. В таких опытах определяется разборчивость речи, передаваемой по тракту связи. Разборчивость определяется как отношение количества правильно принятых сигналов к общему количеству переданных сигналов. Нужно иметь в

виду, что, когда мы говорим о сигналах, используемых в языке, результат правильного опознания (приема) сильно зависит от того, какой именно сигнал передается. Так, если по каналу связи, даже очень «плохому», т. е. с узкой полосой пропускания, с большими помехами, передаются предложения или слова, то они могут узнаваться испытуемыми не из-за собственно акустических характеристик, а благодаря исключительной способности человека по неполному и искаженному акустическому сигналу восстанавливать смысл высказывания на знакомом ему языке. Поэтому для оценки акустических характеристик вторичного сигнала используют бессмысленные звукосочетания — слоги или квазислова. Важной фонетической задачей является определение статистических характеристик звуковых последовательностей, свойственных данному языку: ведь используемые при измерении разборчивости слоги и квазислова должны быть организованы по тем же статистическим закономерностям, что и нормальные слоги и слова. Эта прикладная задача вызвала к жизни большое количество статистических исследований на материале самых разных языков и не только обеспечила базу для измерения разборчивости речи, но и имела большое значение для типологических исследований.

ОСНОВНЫЕ АКУСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ЗВУКОВ РЕЧИ

Звук речи (как любой звук вообще) является результатом воздействия колебательных движений воздушной среды на слуховой аппарат человека. Эти колебательные движения могут быть охарактеризованы по следующим основным признакам.

1. Частота колебательных движений определяется числом колебательных движений в единицу времени. Единицей измерения частоты является герц (в честь немецкого физика Герца): когда говорят, что частота какого-то звука 100 Гц, это означает, что, если бы этот звук длился одну секунду, мы могли бы зарегистрировать за это время 100 колебательных движений. От частоты звука зависит ощущение его высоты: чем больше частота, тем более высоким является звук для слуховой системы. Единицей высоты звука является мел. Отношение между частотой и высотой звука можно увидеть на рис. 1. В специальной литературе частоту обозначают латинской буквой f (англ. frequency).

2. Интенсивность, или сила, звука определяется амплитудой колебательных движений. Звуковые колебания, передаваемые воздушной средой, можно представить себе как последовательные сгущения и разрежения воздуха, т. е. изменения воздушного давления по сравнению с атмосферным. Это давление измеряется силой, действующей на единицу площади. Громкость звука, зависящая от его интенсивности, возрастает с увеличением звукового давления. Порогом слышимости называется

такое значение звукового давления, при котором звук еле слышится. При постепенном увеличении звукового давления громкость звука возрастает до тех пор, пока не перерастает в ощущение боли. Это значение звукового давления называют порогом болевого ощущения. Разница значений звукового давления

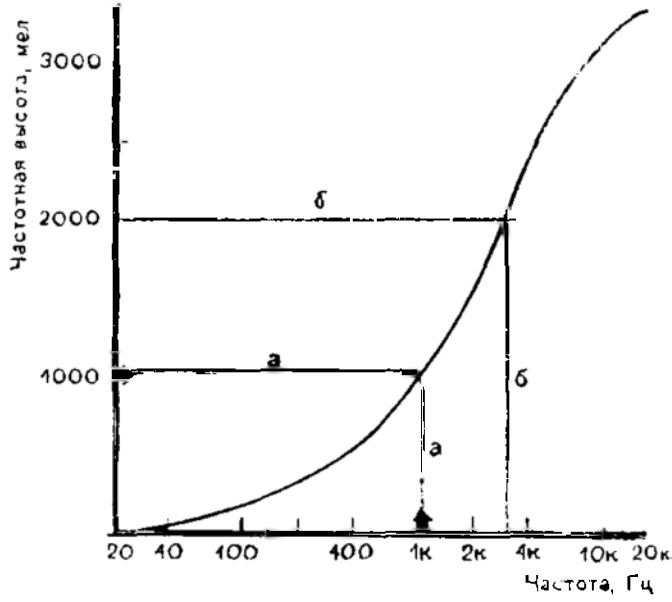


Рис. 1.

на пороге слышимости и на пороге болевого ощущения очень велика, поэтому в акустике при измерении силы звука используют логарифмические единицы — децибелы (сокращенное обозначение — дБ, в честь известного физика Белла). С помощью этой единицы выражается уровень силы звука по отношению к некоторым условным значениям, называемым стандартными нулевыми уровнями. Например, для тона с частотой 1000 Гц уровень силы звука на пороге слышимости равен 0 дБ, а на пороге болевого ощущения — 130 дБ. Чувствительность уха к разным частотам различна, поэтому восприятие громкости звука зависит от его частоты. В акустике громкость любого звука измеряют величиной уровня звукового давления равногромкого с данным звуком тона с частотой 1000 Гц. Эта величина называется уровнем громкости и измеряется в децибелах. На рис. 2 приведены кривые равной громкости для звуков с различной частотой. Видно, что и высокие — выше 2000 Гц, и низкие — ниже 500 Гц звуки требуют гораздо большей интенсивности по

сравнению с остальными, чтобы их громкость воспринималась как равная громкости тона с частотой 1000 Гц.

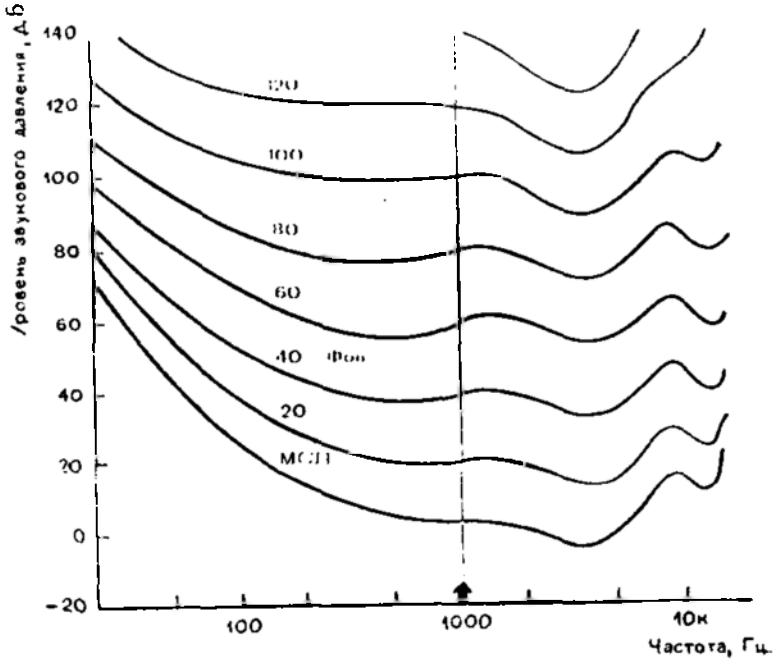


Рис. 2.

3. Временны́е характеристики звука. Для речевых звуковых сигналов существенно, что они имеют разное время звучания. *Это связано как с артикуляторными свойствами* сегментных единиц, так и с разворачиванием во времени всего сообщения, где временны́е различия несут различительную функцию. Длительность звуков речи и их последовательностей в высказывании измеряется тысячными долями секунды — миллисекундами (сокращенное обозначение — мс). Пределы изменения длительности очень велики — например, длительность взрыва губно-губного взрывного глухого согласного [р] — около 15 мс, длительность ударного гласного — от 150 до 200 мс, а длительность высказывания не ограничена никакими фонетическими пределами и зависит лишь от размера реализуемого текста.

4. Спектральные характеристики и качество звука. Акустические свойства звуков зависят от того, как образуются эти звуки в речевом тракте. Речевой тракт может быть рассмотрен как система, обеспечивающая взаимодействие источни-

ков звука и резонансных свойств надгортанных полостей речевого тракта. В качестве источника выступают голосовые связки, колебательные движения которых создают низкочастотный квазипериодический гармонический колебательный процесс. Этот источник звука участвует в образовании гласных, сонантов и звонких шумных согласных. При образовании взрывных согласных источником звука является процесс выравнивания воздушного давления за местом смычки (которое всегда выше, чем атмосферное) и атмосферного давления. При этом возникает так называемый импульсный шум, характеризующийся небольшой длительностью и нерегулярным шумом, расположенным во всем частотном диапазоне. При образовании щелевых согласных выступает второй вид шумового источника — турбулентный, образующийся в результате трения воздушной струи, проходящей по речевому тракту, о края щели, образуемой при артикуляции того или иного согласного.

Резонансная система речевого тракта образуется из надгортанных полостей — полости глотки, ротовой и носовой полости. Полость носа может быть включена в резонансную систему — в случае, когда проход воздушной струи в нее открыт, что бывает при опущенном мягком небе, или выключена — при поднятом мягком небе. Полость рта и полость глотки меняют свою конфигурацию, размер и величину выходного отверстия в результате движений языка и губ. Эти изменения определяют резонансные свойства речевого тракта, так что при артикуляции конкретных звуков источник вызывает колебательные движения в такой резонансной системе, свойства которой зависят от характера артикуляции.

С акустической точки зрения в результате взаимодействия источника (или нескольких источников) с резонансными частотами возникает сложный звук, восприятие которого зависит от нескольких акустических параметров.

Основные методы анализа акустических характеристик. Особенностью акустических сигналов, возникающих при артикуляции звуков речи, являются чрезвычайно малые значения, характеризующие изменения звукового давления. В связи с этим зарегистрировать непосредственно колебательные движения во время звучания практически невозможно. Как уже было сказано выше, исследователи речи чаще всего имеют дело не с первичными, а со вторичными сигналами.

Существуют два основных вида регистрации акустических явлений. Во-первых, это регистрация полной огибающей звуковой волны при помощи приборов, называемых осциллографами. Во-вторых, это регистрация акустических параметров после предварительного анализа — этот анализ производится несколькими способами при помощи спектроанализаторов. Спектральный анализ звука предполагает установление того, какие частоты участвуют в образовании данного звука и какова их

интенсивность по отношению друг к другу. Для такого анализа используют метод фильтрации: вторичный сигнал поступает на гребенку фильтров или проходит через фильтр с изменяющейся частотой. При этом на выходе фильтра возникает сигнал в том случае, если в анализируемом звуке содержится колебание, совпадающее по частоте с собственной частотой фильтра. Интенсивность отклика коррелирует с интенсивностью соответствующей частотной составляющей в анализируемом звуке. В результате такого анализа получают так называемые амплитудно-частотные спектры, или мгновенные значения спектра. Такой анализ может дать и представление об изменении спектра звука во времени, если получать такие мгновенные значения для следующих друг за другом коротких частей одного звука.

Для более быстрого анализа изменений спектральных характеристик звуков во времени используют специальные приборы — динамические спектрографы. Ось частот на динамической спектрограмме расположена по вертикали, а горизонтальная ось соответствует оси времени. Значения интенсивности частотных составляющих коррелируют с яркостью свечения точек, соответствующих выходным значениям фильтров. При съемке спектрограмм на светочувствительную пленку или бумагу более яркие свечения дают бо́льшую степень затемненности, что позволяет визуально оценивать относительную интенсивность составляющих спектра. До недавнего времени и осциллографический, и спектральный анализ речи производился при помощи так называемых аналоговых систем. В последние годы, в результате развития вычислительной техники, повсеместное распространение приобрел цифровой анализ, дающий возможность наиболее совершенного акустического анализа.

Для анализа речевого сигнала используют специальные аналого-цифровые преобразователи, позволяющие каждый звук или звуковую последовательность представить в виде набора целых чисел и записать в памяти компьютера именно как набор чисел. С помощью специальных программ производят регистрацию и анализ акустических параметров в тех видах, которые ранее производились осциллографами и спектрографами. Программным способом достигается и возможность модификации записанных в память компьютера звуков — ослабление или усиление тех или иных частотных составляющих, сегментация на более короткие отрезки и т. д. Это существенно для изучения значимости тех или иных акустических свойств звуков для их восприятия. Возможность такого анализа обеспечивается использованием другого преобразователя — цифро-аналогового, при помощи которого записанный в виде набора чисел звук может быть воспроизведен как аналоговый акустический сигнал. На рис. 3 приведены осциллограммы и спектрограммы слога *за*, полученные на аналоговой аппаратуре (рис. 3, *а*) и при помощи цифрового анализа программным способом (рис. 3, *б*).

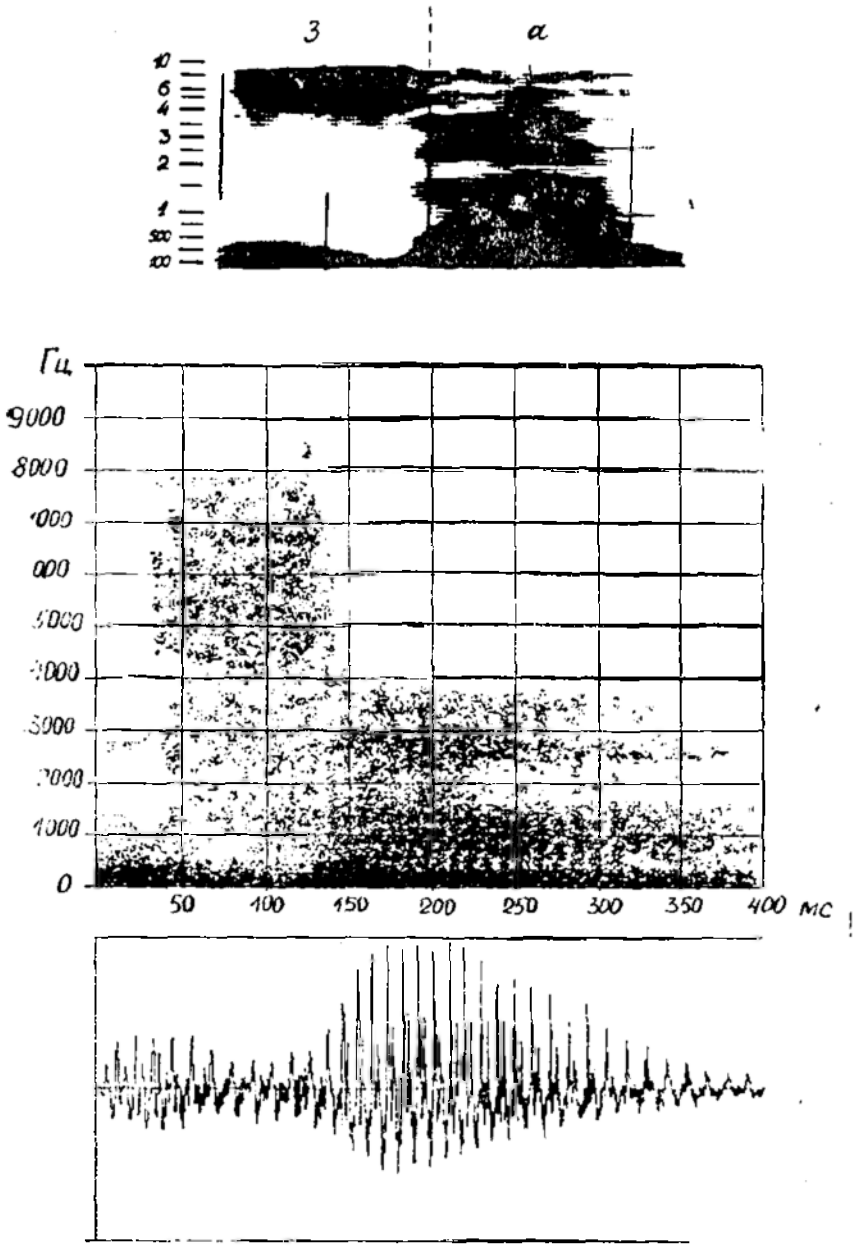

3
a
10
6
4
3
2
1
500
100
Гц
9000
8000
0
50
100
150
200
250
300
350
400
мс

Рис. 3.

Специальным видом анализа акустических свойств речевого сигнала является анализ изменений частоты голосового источника. Это необходимо при исследовании интонации, поскольку основной компонент интонации — мелодика — зависит именно от характера изменений частоты колебательных движений, совершаемых голосовыми связками. Осциллографическая кривая дает достаточно сведений для измерения величины периодов основного тона, и большое количество экспериментально-фонетических исследований интонации основано на таких измерениях. При работе с осциллограммой измеряют или величину каждого периода, что чрезвычайно трудоемко, или длительность группы периодов, получая таким образом усредненное представление об интересующих изменениях. Известны также и специальные приборы, обеспечивающие автоматическое выделение и измерение периодов колебания голосовых связок, — интонографы. В настоящее время такой анализ производится также с помощью компьютера программным способом.

Гласные и согласные на динамических спектрограммах. Динамические спектрограммы очень широко используются в работах по прикладной фонетике. Для правильного понимания того, как могут интерпретироваться эти спектрограммы, рассмотрим более подробно их особенности на примере русских гласных и согласных.

Различия между гласными определяются тем, что при их произношении различаются надгортанные резонаторы. В тех полостях речевого тракта, которые выступают как резонаторы, возбуждаются колебательные процессы, при этом возникает сложный в акустическом отношении звук, для характеристики которого важно определить, какие области спектра в нем усилены.

Области усиления в спектре гласного, зависящие от резонансных характеристик речевого тракта, называют формантами и обозначают латинской буквой *F*. Для достаточно полной характеристики гласного, т. е. такой характеристики, которая позволяет отличить данный гласный от всех остальных, достаточно учитывать частоты первых двух формант — *FI* и *FII*. Для гласных с высокой частотой *FII* существенны и характеристики *FIII*.

Между формантными частотами и артикуляционными признаками гласных существует зависимость. Первая форманта связана с таким признаком гласного, как степень подъема: чем более открытым является гласный, тем выше частота первой форманты. Вторая форманта связана с рядом гласного: самый передний гласный имеет самую высокую частоту второй форманты, самый задний — наиболее низкую. Огубленность понижает частоты всех формант. Эта общефонетическая зависимость хорошо видна на рис. 4, где приведена диаграмма соотношения частот формант и артикуляционных признаков гласного.

Каждый гласный на динамической спектрограмме представ-

лен как изменяющаяся акустическая картина: в идеальном случае стационарному участку гласного — т. е. такому участку, на котором частоты формант не изменяются, — предшествует первый переходный участок. Свойства этого участка зависят от характеристик предшествующего звука. После стационарного

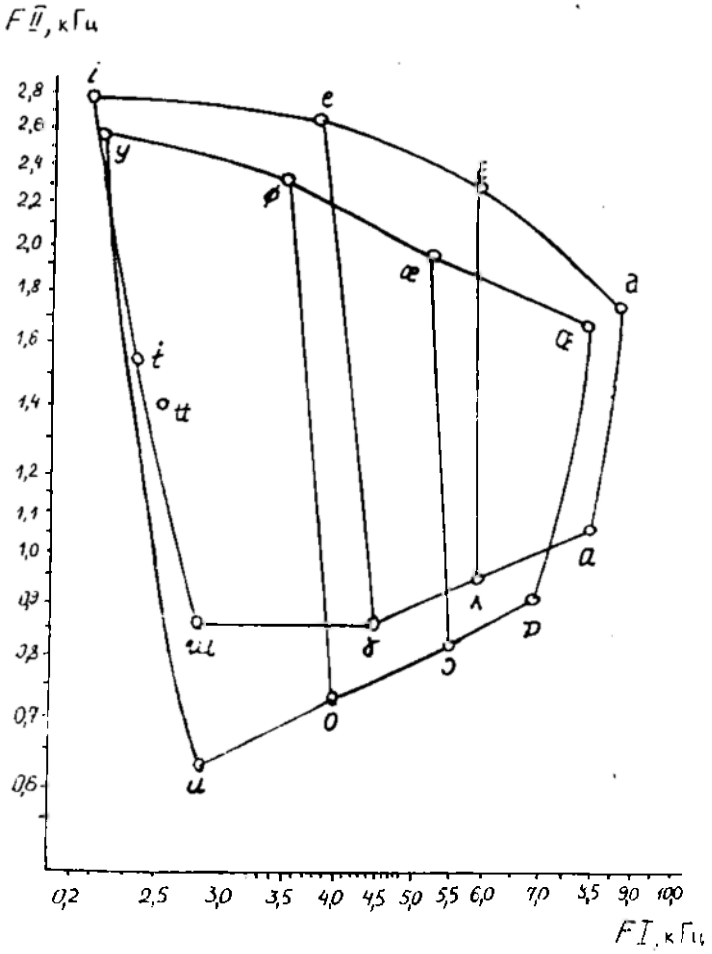


Рис. 4.

участка следует второй переходный участок, также зависящий от следующего согласного. Возможны случаи, когда один или два из этих участков отсутствуют. На рис. 5 приведены динамические спектрограммы, иллюстрирующие эти возможности. Наиболее общими признаками, характеризующими все гласные, яв-

ляются четкая формантная структура и большая интенсивность звука. Оба эти свойства определяются наличием голосового источника и активной ролью резонансной системы.

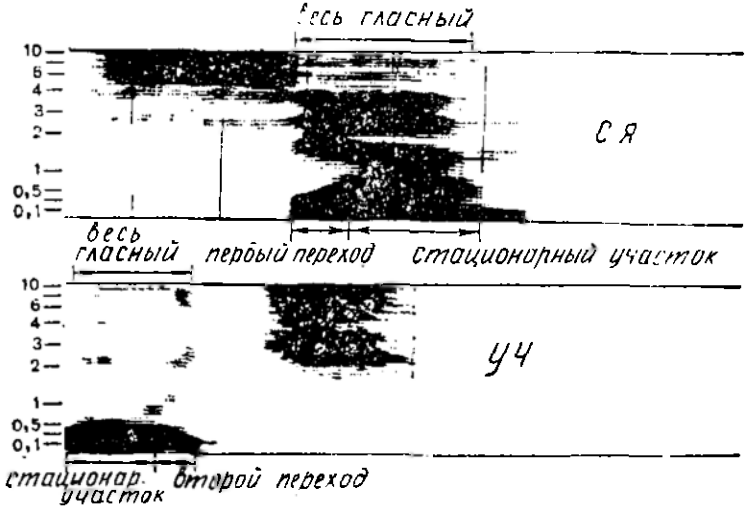


Рис. 5.

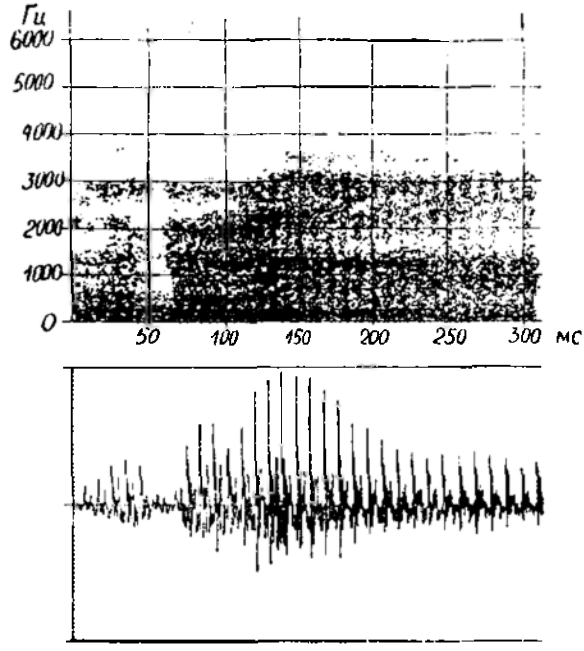
Динамические спектрограммы согласных несут информацию о свойствах источников, участвующих в их образовании, а также о месте их расположения в речевом тракте. На рис. 6 приведены динамические спектрограммы разных согласных и даны объяснения их особенностей (*а* — голосовой источник, *б* — импульсный источник, *в* — турбулентный источник).

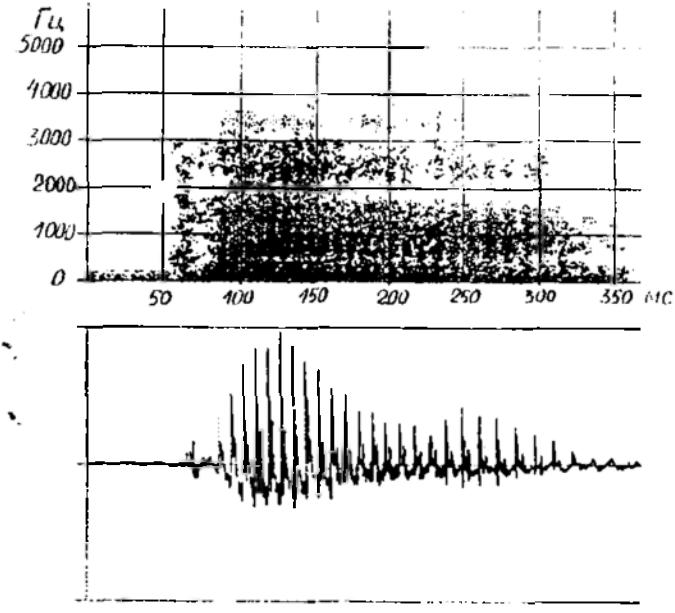
Из краткого описания основных акустических свойств гласных и согласных можно сделать вывод о том, что эти свойства могут быть использованы для автоматического распознавания речи.

Фонетические проблемы автоматического распознавания. Для проблемы автоматического распознавания существенны следующие свойства распознающей системы: количество распознаваемых единиц, ограничения, связанные с голосом диктора, свойства распознаваемого сообщения, время распознавания, условия распознавания.

Рассмотрим более подробно, какие проблемы возникают в каждом конкретном случае.

1. Количество распознаваемых единиц. Если ставится задача распознавать небольшой набор команд или цифр — а это довольно распространенная в технике задача, — целесообразно анализировать целостный акустический образ каждого слова и вырабатывать такие признаки распознавания,


Гц
6000
5000
4000
3000
2000
1000
0
50
100
150
200
250
300
мс


Гц
5000
4000
3000
2000
1000
0
50
100
150
200
250
300
350
мс

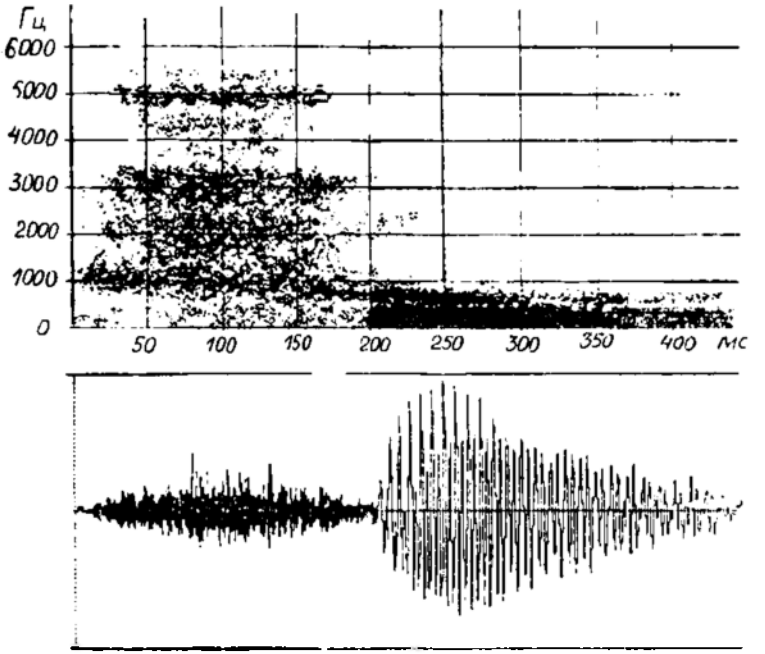


Рис. 6.

которые были бы и надежными и экономными. Например, при распознавании названий чисел от одного до девяти такими признаками может быть число слогов, место ударного гласного и его качество, а также наличие или отсутствие конечного согласного. Эти признаки пригодны для распознавания именно этого набора слов и не требуют большого объема памяти распознающего устройства. В том случае, когда необходимо распознавать достаточно большой набор слов, такой подход неприемлем, поскольку гораздо экономнее распознавать каждое слово как последовательность минимальных сегментных звуковых единиц. Важная задача для фонетиста — определение количества этих единиц. Ясно, что оно не может быть меньше количества фонем, однако в целом ряде распознающих систем используют большее число — в этом случае их называют «фонемами распознавания». От количества единиц, используемых при распознавании, зависит выбор полезных признаков, по которым различаются эти единицы.

2. Выбор диктора. Распознающая система может быть предназначена для распознавания команд, произнесенных одним голосом, — такая задача является наиболее простой, по-

скольку вариативность акустических свойств речевого сигнала очень сильно зависит от качества голоса. В том случае, когда ставится задача распознавания любого голоса, решение может быть двояким: устройство может работать с предварительной настройкой на голос конкретного диктора или же создается такая система, которая должна распознавать речевой сигнал при любых индивидуальных особенностях. В связи с тем, что индивидуальные особенности голоса сильно влияют на акустические свойства речевых сигналов, возникла мысль о возможности по акустической картине определять принадлежность высказывания конкретному диктору.

3. Свойства распознаваемого сообщения. Сложные распознающие системы предполагают возможность работы со связными текстами. Эта задача гораздо сложнее, чем распознавание большого словаря. Акустические свойства звуковых элементов, образующих речевое сообщение, подвергаются значительным модификациям в результате взаимодействия сегментных и супрасегментных характеристик. Кроме того, в результате смысловой избыточности, характерной для всякого высказывания, возрастает фонетическая неопределенность, приводящая к известному размыванию собственно фонетических характеристик, в том числе и акустических. Как показывает анализ развития таких распознающих систем, с усложнением задачи увеличивается доля дополнительной информации, носящей собственно лингвистический характер, или вводятся некоторые ограничения, позволяющие избежать возможной неопределенности решения. Примером дополнительной информации, вводимой в системы, распознающие связный текст, является для текстов на русском языке такая информация, как правила фонемной интерпретации фонетически неопределенных гласных или соотношение фонетической и фонемной сегментации заударных частей словоформы. Примером вводимых ограничений являются ограничения на порядок слов в предложении или ограничения в темпе речи.

Создание системы, распознающей неограниченный словарь на данном языке, должно базироваться на сложных лингвистических правилах, учитывающих не только акустическую вариативность речевого сообщения, но и статистические свойства звуковых цепей и закономерности образования фонетической целостности высказывания.

4. Время распознавания. В естественном речевом общении распознавание происходит в реальном масштабе времени, и создатели автоматических систем стремятся к обеспечению текущего распознавания акустического сигнала. Такое обеспечение требует решения двух кардинальных задач: задачи сегментации непрерывного по своей акустической природе высказывания на минимальные единицы, соотносимые с «фонемами распознавания», и задачи отождествления этих сегментов с эталон-

ными образцами, хранящимися в памяти распознающего устройства. Обе эти задачи в свое время явились полной неожиданностью для специалистов в области речевых технологий, поскольку и сегментация, и отождествление считались очень простыми операциями чисто акустического порядка. Углубленное рассмотрение показало, что пользующийся звуковой речью человек осуществляет сегментацию и отождествление не на акустических, а на лингвистических основаниях. Это означает, что при автоматическом распознавании следует идти двумя путями. Первый путь — моделируя речевую деятельность человека, привлекать к решению обеих задач правила языка, т. е. уже на уровне перехода от акустической картины к предварительной «транскрипции» текста вводить сложные лингвистические операции; второй путь — это отказ от моделирования речевой деятельности человека и создание таких алгоритмов интерпретации акустического сигнала, которые ориентированы только на возможности распознающего устройства.

5. Условия распознавания. Передача речи по трактам связи — в том числе и в случае автоматического распознавания — может происходить в разных условиях, и часто требуется специальное выделение речевого сигнала из того шумового фона, на котором он передается. Эта задача вызвала к жизни специальное направление исследования акустических характеристик сигналов, которое носит название «речь — не речь». Обнаружение речевого сигнала на фоне шума также требует моделирования соответствующих процессов в речевой деятельности человека; выявление признаков конкретных звуков в условиях зашумления — одна из труднейших проблем автоматического распознавания.

Синтез речи в исследовательских и прикладных программах. Создание речеподобных звуков без участия человека привлекало к себе уже очень давно, и эта задача, даже если первоначально она была чисто рефлексивной, сыграла большую роль в изучении фонетических свойств языковых единиц. Синтезирующие устройства, функционирующие в настоящее время, можно расклассифицировать в зависимости от тех целей, которые ставятся при их создании, и от тех способов, которыми получается синтезированная речь.

Синтезированная речь широко используется в современной науке для изучения механизмов восприятия речи человеком, для моделирования тех психофизиологических процессов, которые делают столь совершенным речевое общение в человеческом обществе. Достоинство синтезированных сигналов, используемых в экспериментах такого рода, заключается в том, что исследователь может изменять характеристики этих сигналов именно в тех направлениях, которые соответствуют исходной гипотезе, и проверять значимость этих изменений для испытуемых, воспринимающих синтезированные сигналы.

Примером работы такого рода может служить исследование Н. А. Федоровой, посвященное изучению полезных признаков словесного ударения в русском языке. В качестве экспериментального материала использовались синтезированные стимулы длительностью 400 мс, с частотами первых двух формант, соответствующими частотам естественного гласного [а]; эти стимулы содержали короткую паузу, благодаря чему они воспринимались как звуковая цепочка [ара]. Передвигая место этой паузы по стимулу, Н. А. Федорова меняла длительность и первого и второго гласного, что приводило к восприятию как ударного того гласного, который имел бóльшую длительность. Эти результаты хорошо соответствовали полученным ранее данным об особой роли длительности как ведущего фонетического коррелята словесного ударения в русском языке. Поскольку исследователь пользовался синтезированными стимулами, оказалось возможным одновременно с длительностью гласных изменять их интенсивность и высоту и на основе статистического анализа результатов опознания определить значимость этих параметров для определения места ударения в экспериментальным стимуле. Оказалось, что именно длительность является тем главным признаком, который определяет ритмическую структуру синтезированного стимула.

Другим примером использования синтезированных стимулов при изучении механизмов речевой деятельности человека является использование множества четырехформантных гласных для определения критических значений частот формант при переходе от одного качества гласного к другому при их восприятии.

Практическое применение синтезированной речи связано с созданием наиболее комфортных условий общения человека с разнообразными техническими устройствами. Так, в последние годы довольно широкое распространение получили системы речевого общения водителя автомобиля с управляющим компьютером, который во время пользования автомобилем сообщает о таких важных показателях, как соответствие выбранной скорости качеству дорожного покрытия, количество бензина, температура в охлаждающей системе и т. д. Если водитель забыл закрыть дверь салона или крышку багажника, компьютер сообщает ему об этом и не разрешает начать движение.

Как уже было сказано выше, большое распространение находит идея синтеза речевого сигнала по письменному тексту. Такие системы предназначены для использования в гуманитарных целях — для людей с ослабленным зрением или вовсе слепых. Синтез «текст — речь» позволяет озвучивать любые тексты, т. е. делать доступными эти тексты и при обучении незрячих людей, и при знакомстве с памятниками литературы всех времен и народов. При создании таких систем возникает одна из промежуточных задач, решение которой определяется и фонети-

ческими, и техническими факторами: это задача транскрибирования орфографического теста в последовательность фонетических символов. Для решения этой задачи разрабатываются системы автоматического транскрибирования, позволяющие переходить к фонетической записи любого текста на основе использования специальных компьютерных программ. Эти программы могут требовать предварительной расстановки ударения в словах, что, естественно, делает транскрибирование не полностью автоматическим. При использовании более сложных лингвистических правил возможно создание транскрипторов, не требующих предварительной ручной обработки текстов.

Создание синтезированной речи хорошего качества, т. е. звучащей и естественно, и разборчиво, требует соотнесения знаков фонетической транскрипции с соответствующими акустическими параметрами аллофонов, обозначаемых этими знаками. Накопление сведений такого рода — одна из важнейших задач прикладной фонетики. Другим условием для создания высококачественного синтеза является алгоритмизация правил интонационного оформления высказывания. Эта задача стимулирует исследования в области интонации связной речи.

Рассмотрим теперь те способы, которые обеспечивают получение речеподобных синтезированных сигналов без участия человека как источника этих сигналов. Условно эти способы можно разбить на две группы: технический синтез и компилятивный синтез. При техническом синтезе речеподобные сигналы создаются или с помощью технических устройств, или программным способом. При компилятивном синтезе используются естественные речевые звуки, из которых склеиваются слова, фразы и целые тексты.

Технический синтез может быть охарактеризован как моделирование разных сторон речепроизводства. В истории развития синтеза можно выделить несколько этапов.

В конце XVIII в. была создана первая «говорящая машина», представлявшая собой акустические резонаторы, по форме аналогичные речевому тракту человека, а аналогом голосовых связок являлись вибрирующие язычки. В дальнейшем было построено несколько таких говорящих машин, принципом работы которых было повторение устройства речевого тракта человека и его работы во время производства речи. Этот период развития технического синтеза можно назвать механическим. С развитием электроакустики возникли новые подходы к проблеме и были созданы электрические синтезаторы. Первые синтезаторы этого типа также ориентированы на моделирование артикуляторных движений во время речи, т. е. отдельные элементы системы как бы «соответствовали» определенным участкам речевого тракта. В настоящее время синтезаторы — аналоги речевого тракта используются в научных исследованиях для определения роли отдельных участков тракта в образова-

нии акустических свойств речевого сигнала. Другой разновидностью синтезаторов являются такие приборы, которые синтезируют речеподобные звуки на основе собственно акустических свойств этих звуков. Синтезируется или общая спектрально-временная картина, повторяющая свойства естественных звуков, или формантная картина, однако во всех случаях учитываются характеристики источника, голосового тракта и временной организации акустического процесса. Развитие вычислительной техники сделало возможным использование ЭВМ как синтезирующих устройств. Способность ЭВМ хранить большое количество информации и совершать логические операции с исключительно большой скоростью позволяет использовать их и для управления аналоговыми голосовыми трактами, и для формантного синтеза, и для синтеза спектра.

Компилятивный синтез как реальный в техническом отношении метод возник в связи с развитием вычислительной техники. Современный компьютер позволяет преобразовывать аналоговый речевой сигнал в цифровую форму и хранить его в памяти, а также производить разнообразные модификации и анализ множества хранящихся сигналов. Таким образом исследователь может получить набор естественных сегментов, необходимых для перехода от знаков фонетической транскрипции к реальному звучанию. Основная задача при разработке систем компилятивного синтеза определяется необходимостью создать минимальный набор элементов при их совместимости друг с другом. Не менее важная задача — формирование естественной просодики. Как показали первые опыты, компилятивный синтез должен базироваться на использовании элементов разной протяженности — от звуковой цепочки, соответствующей основе слова, до переходного участка гласного. В настоящее время компилятивный синтез широко применяется в обучающих программах при изучении иностранных языков — в том числе и русского как иностранного — и при обучении людей с нарушением зрения.

ОСНОВНЫЕ ПРОБЛЕМЫ ИССЛЕДОВАНИЯ АРТИКУЛЯЦИИ

Произносительная деятельность человека представляет собой чрезвычайно сложный процесс, в котором непосредственно наблюдаемые артикуляции являются лишь конечным результатом иерархически организованной процедуры перехода от семантической программы планируемого высказывания к ее развертыванию в виде последовательностей команд, необходимых для осуществления этого высказывания произносительными органами. Как мы уже видели, правильное представление об артикуляторных процессах необходимо при создании систем синтеза речи. Особое значение приобрели работы, посвященные

артикуляторным процессам, в связи с теми трудностями, которые обнаружили создатели систем автоматического распознавания речи. В начале 60-х годов нашего века возникла концепция, согласно которой именно моторная деятельность человека обеспечивает оптимальную обработку акустических сигналов, представляющих звуки речи. Более того, была высказана мысль, что с физиологической и с фонетической точек зрения можно допустить, что фонемы как смыслоразличительные элементы существуют только на уровне синтеза артикуляторной программы, соответствуя некоторым элементарным указаниям, относящимся к структуре комплексов артикуляторных движений, входящих в эту программу. Стремление разобраться в структуре артикуляционного процесса привело к необходимости разработать такие методы анализа артикуляций, которые соответствовали бы поставленной задаче.

МЕТОДЫ ИССЛЕДОВАНИЯ АРТИКУЛЯЦИИ

Экспериментальная фонетика возникла как наука о физиологии звуков речи, и первые методики регистрации были направлены именно на важнейшие артикуляционные признаки: речевое дыхание, положение произносительных органов, поведение голосовых связок и т. д. Значительное место в ранних работах занимало и так называемое самонаблюдение, когда экспериментатор путем анализа мускульных ощущений во время произнесения того или иного звука пытался установить его основные признаки. Как правило, первые наблюдения были посвящены описанию положения произносительных органов, т. е. были статическими. Представление о развертывании артикуляции во времени исчерпывалось учением о трех фазах артикуляции — приступе, выдержке, отступе; при этом наиболее важной считалась, конечно, фаза выдержки. Инструментальный анализ артикуляций в первый период развития физиологической фонетики базировался на следующих данных: фотосъемки губных артикуляций, палатография и рентгенография.

Фотосъемки (а в дальнейшем и киносъемки) губных артикуляций — надежный способ регистрации такого важного артикуляционного признака, как положение или движение губ во время произнесения гласных и некоторых согласных. При детальном анализе таких материалов оказалось, что губные артикуляции позволяют опознавать не только губные, но и другие звуки (чтение «по губам» — существеннейший аспект обучения слабослышащих и глухих, а также и восприятия речи в специальных условиях при дефиците звуковой информации).

Палатография — это метод, позволяющий определять области касания языка с твердым небом во время артикуляции. Для этого изготавливается искусственное твердое небо — тоненькая пластинка, повторяющая конфигурацию собственного

неба испытуемого. Перед началом регистрации искусственное небо покрывается легко смывающимся составом или порошком, например тальком, и вставляется в рот испытуемого. После произнесения звука пластинка осторожно вынимается изо рта, и получившийся отпечаток языка на небе срисовывается или фотографируется. В настоящее время этот метод несколько упрощен: палатография производится без изготовления искусственного неба — собственное небо испытуемого окрашивается, и после артикуляции соответствующего звука фотографируется отражение этого неба на специальном зеркале, помещаемом в ротовую полость.

Для получения информации о положении языка во время артикуляции использовался метод рентгенографии. По средней линии языка проводится полоска составом, который не пропускает рентгеновских лучей, и таким образом на рентгенограмме можно наблюдать конфигурацию языка во время произнесения звука. В дальнейшем в экспериментальной фонетике стали применять метод кинорентгенографирования, который обеспечивал получение сведений о динамике артикуляторного процесса. Однако в настоящее время оба этих способа не имеют сколько-нибудь широкого распространения в связи с опасностью длительного облучения испытуемого, сложностью аппаратуры и трудоемкостью обработки графического материала. Одним из образцовых исследований динамики речеобразования являются исследования Л. Г. Скалозуб. В ее работах читатель найдет много доказательств чрезвычайной трудоемкости систематического анализа артикуляторных движений.

В начале 60-х годов нашего века была создана уникальная и в определенной мере универсальная система комплексной регистрации артикуляторных параметров речи. Эта система включает в себя целый набор датчиков, электронные схемы для усиления и предварительной обработки сигналов, поступающих с датчиков, схемы логической обработки информации, регистраторы и автоматические измерители. Одновременно и непрерывно можно получать в виде электрических сигналов следующие показатели: внутриротовое давление, скорость движения воздуха из ротового и из носовых давлений при фонации, толчки воздуха, выходящего из этих отверстий при фонации, дыхательные движения на разных уровнях груди и живота, движения нижней челюсти, смыкание и размыкание губ, явление огубленности, динамическую палатограмму (с возможностью одновременно регистрировать динамику касания языком 10 и более раздельных зон твердого неба). Вместе с регистрацией этих артикуляторных параметров система дает возможность записывать нормальный звуковой сигнал, являющийся результатом этих артикуляций.

Артикуляторная организация элементов высказывания. Использование совершенных методов анализа артикуляций позво-

лило по-новому взглянуть на некоторые важные вопросы теоретической и прикладной фонетики. Прежде всего, появилась возможность проанализировать динамику произнесения конкретных звуков. Оказалось, что классическое представление о трехфазной организации временной программы каждого звука не соответствует действительности. В непрерывной звуковой цепи отступ первого элемента совпадает с приступом следующего за ним, однако отклонения от классической схемы этим не исчерпываются: в естественном речевом потоке очень часто отсутствует и так называемая фаза выдержки. Это обстоятельство приводит к предположению о том, что произносительная программа не связана с реализацией каждого отдельного звука. Специальные эксперименты показали, что минимальной произносительной единицей, которая выполняется в результате одной произносительной программы, является открытый слог. В непрерывной речевой цепи артикуляторные движения построены по достаточно простому алгоритму: речевая цепь — это последовательность слогов, а каждый слог кончается с завершением гласного, независимо от количества и качества предшествующих согласных. Морфемные и межсловные границы не влияют на реализацию этого алгоритма.

Интерес к механизмам управления артикуляторными движениями привлекает внимание и к тем сведениям о работе мозга, которые обсуждаются при исследовании функциональной асимметрии левого и правого полушарий коры головного мозга.

Огромную область исследований представляют собой проблемы нарушений речи, вызванных различными мозговыми поражениями. В настоящее время существуют достаточно точные методики диагностики этих поражений по характеру артикуляторных нарушений и разрабатываются специфические способы реабилитации речевой функции в зависимости от вида мозговых поражений.

Очень большое внимание уделяется тем нарушениям артикуляции, которые наблюдаются у детей. Логопеды создали чрезвычайно подробную классификацию нарушений устной речи ребенка, и при этом они соотносят каждый из видов с определенными причинами и возможностями воздействия на эти нарушения.

Нарушения устной речи делятся на два типа: 1) фонационного (внешнего) оформления высказывания — нарушения собственно произносительной стороны; 2) нарушения структурно-семантического (внутреннего) оформления высказывания — системные или полиморфные нарушения речи. В свою очередь, расстройства фонационного оформления зависят от нарушения следующих звеньев: голосообразования, темпо-ритмической организации, интонационно-мелодической и звукопроизносительной. Эти расстройства могут выступать как в чистом виде, так и в комбинациях друг с другом. Раскроем основные термины,

которые употребляются логопедами и которые должен понимать специалист в области прикладной фонетики.

Дисфония или афония — отсутствие или расстройство фонации вследствие патологических изменений голосового аппарата; брадилалия — патологически замедленный темп речи; тахилалия — патологически ускоренный темп речи; заикание — нарушение темпо-ритмической организации речи, обусловленное судорожным состоянием мышц речевого аппарата; дислалия — нарушение звукопроизношения при нормальном слухе и сохранной иннервации речевого аппарата; ринолалия — нарушения тембра голоса и звукопроизношения, обусловленные анатомо-физиологическими дефектами речевого аппарата; дизартрия — нарушение произносительной стороны речи, обусловленное недостаточной иннервацией речевого аппарата.

Нарушения структурно-семантического оформления высказывания представлены двумя видами: алалией — отсутствием или недоразвитием речи вследствие органического поражения речевых зон коры головного мозга и афазией — полной или частичной утратой речи, обусловленной локальными поражениями головного мозга (Логопедия, 1989).

Уже из этого поверхностного перечисления основных видов нарушений речи у детей видно, какое большое значение имеют сведения о соответствующих характеристиках в нормальной речи, и мы знаем, что логопедия широко использует фонетические сведения в работе со своими пациентами.

Среди проблем артикуляторной фонетики значительное место занимают исследования работы голосовых связок. Как мы уже говорили, голосовые связки выполняют роль источника звука при образовании гласных, сонантов и звонких согласных. Кроме того, их работа обеспечивает мелодический компонент интонации, являющийся важнейшим интонационным средством. Теория голосообразования развивается специалистами в области вокальной речи, в области детской речи, логопедами и другими исследователями. Основная проблема, являющаяся дискуссионной, заключается в том, можно ли рассматривать колебательный процесс во время фонации как пассивный, при котором колебания возникают в результате прохождения тока воздуха через сомкнутые края связок, а частота колебаний зависит от степени эластичности голосовых связок, — или же голосовые связки человека не пассивно колеблются под действием тока воздуха, а активно сокращаются под влиянием управляющих нервных импульсов.

В свете акустической теории речеобразования сведения об артикуляторных процессах приобретают особое значение для обработки акустических свойств речевых сообщений.

ИССЛЕДОВАНИЯ ВОСПРИЯТИЯ РЕЧЕВЫХ СООБЩЕНИЙ

Слуховая система человека обеспечивает преобразование воздушных колебаний в нервные импульсы, свойства которых используются для фонетической интерпретации и принятия фонемных решений в коре головного мозга. Поскольку слуховая обработка звуков речи является очень совершенным процессом, обеспечивающим обнаружение, распознавание и идентификацию поступающих на вход системы сигналов в реальном масштабе времени, создатели систем распознавания речи стремятся моделировать свойства этого процесса при разработке технических устройств распознавания. При этом обнаруживается, что необходимо исследовать несколько разных ступеней обработки речевых сигналов в слуховой системе человека.

Прежде всего, изучаются те свойства человеческого слуха, которые проявляются при восприятии любых звуков, в том числе и звуков речи. Эти свойства характерны для всех индивидов с нормальным слухом и описываются обычно как соотношение некоторых акустических характеристик, свойственных определенному звуку, с тем, как его оценивает слух человека. К таким описаниям относится уже упоминавшееся соотношение между частотой звука и его высотой, зависимость восприятия громкости звука от его частоты и т. д. Все эти наблюдения относятся к области психоакустики и имеют большое значение не только для создания технических систем, имитирующих работу человеческого слуха, но и для улучшения качества передаваемых по каналам связи речевых сообщений, для создания совершенной звукозаписывающей и воспроизводящей аппаратуры, для улучшения акустических свойств концертных залов, учебных аудиторий и т. д. Психоакустика является частью более широкого раздела науки о восприятии звуков — биоакустики. Благодаря исследованиям биоакустиков мы узнаем о тех особенностях, которые характеризуют слуховую деятельность не только человека, но и других представителей живого мира.

Специальный предмет исследования представляют те особенности слуховой деятельности человека, которые связаны с восприятием именно речевых звуковых сигналов. Несмотря на то, что и некоторые животные других видов способны опознавать и различать звуки речи, только человек сам создает такую звуковую среду, которая обеспечивает возможность обмена смысловой информацией с себе подобными, т. е. только человек порождает и воспринимает сигналы, образующие определенную систему. Известно, что существуют такие универсальные свойства слуховой обработки речевых сигналов, которые присущи любому человеку, независимо от того, на ка-

ком языке он говорит. К универсальным свойствам человеческого слуха относятся такие, как способность различать гласные и согласные, способность извлекать полезную акустическую информацию из тех элементов речевого сообщения, которые не совпадают по времени реализации с распознаваемым звуковым элементом, способность восстанавливать отсутствующий элемент по акустическим свойствам окружающих его звуков, способность учитывать вероятность появления звука и многие другие особенности, сформировавшиеся у человека в процессе развития его мыслительной и речевой деятельности. Изучение этих особенностей восприятия речевых сигналов имеет большое значение для психофизиологии, поскольку затрагивает очень сложные механизмы высшей нервной деятельности человека.

Наконец, особый интерес представляют те свойства восприятия, которые сформированы конкретной звуковой системой языка, являющегося родным для данного индивида. Естественно предположить, что способы слуховой обработки при восприятии речи будут зависеть от количества различающихся звуковых элементов — например, для восприятия речи на русском языке необходимо по некоторым акустическим параметрам идентифицировать не меньше шести разных гласных и тридцати шести согласных, тогда как для восприятия французской — не менее пятнадцати гласных и двадцати согласных. Артикуляторно-акустические свойства гласных и согласных при таких различиях в количестве элементов неизбежно должны сильно различаться, и для прикладной фонетики это один из важнейших вопросов: как изменяются процедуры идентификации звуков при переходе от одной звуковой системы к другой. В первую очередь, это очень важно для такого направления, как обучение иностранным языкам. Возможность обучить правильной артикуляции на неродном языке возникает лишь тогда, когда мы уверены в том, что различия между звуками неродного языка воспринимаются нашим учащимся. Языковая специфика восприятия учитывается и при построении технических систем автоматического распознавания и синтеза речи.

Основные методы исследования восприятия. В опытах по изучению особенностей восприятия существенны такие аспекты методики, как свойства звукового материала, предъявляемого испытуемым, и задачи, которые ставятся перед испытуемыми в эксперименте. Практически исследователя интересует, какие акустические характеристики звуковых единиц важны для носителей языка и для каких именно решений они важны. В связи с этим в качестве стимулов далеко не всегда можно использовать слова, поскольку слово может опознаваться очень хорошо даже при очень сильном искажении его звукового облика. Это обеспечивается стремлением человека в любой звуковой цепочке узнать какое-то слово, а также и тем, что каж-

дое слово обладает дополнительными характеристиками, обеспечивающими его опознание в очень трудных для восприятия условиях.

Стимулы, которые предъявляются в опытах, представляют собой или естественные речевые сигналы разной протяженности, или синтезированные речеподобные сигналы, характеристики которых задаются экспериментатором и могут быть изменены по его желанию и под его контролем. Исходным материалом для получения стимулов первого вида является запись на магнитную ленту речевого материала. Для получения собственно экспериментального материала из исходной записи выделяются и переписываются на чистую магнитную ленту те звуковые сегменты, восприятие которых интересует экспериментатора. Эти сегменты и предъявляются испытуемым. В настоящее время существуют программы, позволяющие производить сегментацию оцифрованных речевых сигналов, хранящихся в памяти компьютера, а также организовывать сам эксперимент в соответствии с исследуемой проблемой. Компьютерная техника позволяет также и создавать синтезированные стимулы с заданными параметрами, и модифицировать естественные и синтезированные стимулы в ходе эксперимента.

Рассмотрим теперь, какие задачи решает испытуемый в этих опытах.

Опознание стимула. В этих опытах испытуемые, *услышав стимул, должны определить, что именно они слышат.* Обычно в таких опытах испытуемые записывают свои ответы в специальных анкетах, и основная трудность — выбор соответствующего знака.

Попарное сравнение. Это специальный метод, позволяющий определить способность испытуемых устанавливать сходство или различие предъявляемых стимулов на основе их акустических характеристик. Сравниваемые стимулы предъявляются парами, и трудность, связанная с выбором знака, здесь отсутствует, поскольку испытуемые лишь констатируют наличие или отсутствие различия.

Метод активного поиска. Это разновидность опыта по попарному сравнению. Как правило, опыты такого рода проводятся на синтезированных звуках, при этом один из звуков имеет постоянные характеристики, и испытуемый получает задание подобрать наиболее похожий или наиболее отличающийся стимул. Испытуемый изменяет характеристики этого стимула, управляя синтезатором, однако сам он не знает, какие именно характеристики он изменяет, и руководствуется только своей слуховой оценкой.

ABX-сравнение. В таких опытах используют три стимула, при этом характеристики двух из них (*A* и *B*) известны, и нужно определить, как оценивается третий, неизвестный (*X*), стимул по отношению к двум первым. Как и в опытах на по-

парное сравнение, испытуемый свободен от необходимости искать знак для стимула, но в отличие от попарного сравнения в этих опытах можно оценивать значение акустических параметров, влияющих на решение.

Имитация. Испытуемым предъявляются стимулы, которые они должны повторить, при этом производится запись произнесенного звука и его акустический анализ, что позволяет судить о том, какие именно признаки являются существенными для испытуемого, когда он имитирует предъявляемый стимул. Исследование имитации важно не только для понимания процессов слуховой обработки речевых сигналов, но и для проникновения в сущность программирования речевой деятельности при порождении высказывания.

Психофизиологическое шкалирование. Шкалирование как метод установления связей между акустическими свойствами сигналов и их оценкой в слуховой системе возникло в рамках классической психоакустики и получило новый импульс для развития в связи с развитием техники речевых исследований. В основе современной теории шкалирования лежат два основных вида отношений: отношение близости и отношение предпочтения. Представление о близости позволяет перейти к понятию субъективного расстояния между элементами, а отношение предпочтения связано с тем, что в ряде ситуаций испытуемый может определить взаимоотношение двух сигналов по некоторому субъективно ясному признаку. На этих общепсихологических предпосылках основано применение шкалирования к исследованию механизмов восприятия звуковых речевых последовательностей.

В конкретных работах, выполненных в рамках теории шкалирования, испытуемые решают следующие задачи относительно предъявляемых речевых стимулов: определение субъективного расстояния между стимулами — сама шкала при этом задается экспериментатором; серийное запоминание стимулов, по результатам которого устанавливается вероятность замен одного стимула на другой и определяется степень их субъективной близости; метод семантически противоположных пар, при котором каждый стимул оценивается по отношению к некоторому набору признаков, задаваемых парами слов противоположного смысла. Испытуемый определяет положение каждого стимула по отношению к крайним значениям признака. Обработка результатов всех опытов по шкалированию предполагает использование сложного статистического аппарата и опирается на математические модели соответствующих операций.

Фонетические аспекты речевого воздействия. Звуковая сторона речи с лингвистической точки зрения характеризуется как форма передачи языковых значений и тем самым как бы не может иметь собственной смысловой нагрузки. Однако известно, что человек как сложная система действует одновре-

менно в разных сферах своих возможностей: тот факт, что носитель языка научился использовать звуковые единицы вне связи с их собственными физическими свойствами, не означает, что эти свойства для него безразличны вовсе. Можно назвать несколько направлений речевой деятельности, в которых первостепенное значение приобретает именно звуковая форма речи.

Прежде всего, это ритмически организованная речь — стихи, сказания, былины, для восприятия которых правильная повторяемость некоторых звуковых признаков играет настолько большую роль, что первоначальный звуковой облик значимых единиц может подвергаться значительным модификациям для сохранения заданной фонетической организации. Это достигается повторением отдельных частей слова, добавлением частиц и разнообразных вставок, растягиванием гласных и т. д. Специальный вид воздействия представляет собой такая фонетическая организация речи, которая по своим характеристикам значительно отличается от обычной для данного языка организации. Оно основано на явлении так называемого фонетического значения, суть которого заключается в том, что человек способен оценивать звуки как самостоятельные явления материального мира и приписывать им определенные оценки. Так, почти для всех людей в паре звуков [а] — [i] первый означает что-то большое, а второй — что-то маленькое, сонанты связаны с «приятными» ассоциациями, а глухие шумные согласные — с неприятными. В обычных текстах вероятность появления звуков определяется статистическими законами, характеризующими звуковую организацию данного языка. Для специального воздействия на слушающего автор прибегает к такой организации текста, при которой появление тех или иных звуков существенно отличается от обычного: если в этом тексте встречается много «приятных» звуков, то и воспринимается он с положительной эмоциональной окраской, если же в нем преобладают «неприятные», то характер восприятия резко меняется.

Зависимость результатов речевого воздействия на слушающих от фонетического оформления речи использовали еще в древности, и в специальных руководствах по риторике мы находим упоминания о необходимости тщательной работы над звучащей речью для достижения нужного эффекта.

Сценическая речь — это поле реализации воздействия фонетических свойств на зрителя. Заметим, что сценическая речь обладает целым рядом особенностей, определяемых ее ролью не только на сцене, но и в обществе в целом, поскольку сценическая речь должна представлять собой воплощение идеальной произносительной нормы, которая является образцом для литературной речи на данном языке. Однако при рассмотрении условий функционирования сценической речи мы обнаружива-

ем целый ряд противоречий между ее идеальными и реальными характеристиками.

Прежде всего, сценическая речь — это имитация речи и изображаемой речевой ситуации, поэтому понятие естественности здесь применимо лишь отчасти. Особая выразительность сценической речи, необходимая для достижения желаемого воздействия, с одной стороны, связана с профессиональными актерскими навыками, с другой стороны — она не может состояться, если в каких-то чертах не отражает стихии обычной, не актерской речи.

Техника актерской речи опирается на такие особенности, как специальное фонационное дыхание, «форсированный» голос и «поставленная» дикция. Будучи хранителем орфоэпической нормы, театр, тем не менее, использует и диалектные фонетические черты для создания речевого портрета персонажа.

Важными прикладными задачами являются такие, как изучение особенностей речи дикторов радио по сравнению с речью дикторов телевидения, особенностей звуковой реализации разных жанров: политические новости, развлекательная программа, спортивный репортаж и т. д.

Большая роль средств массовой информации в жизни общества ставит перед исследователями очень серьезную задачу — активного разъяснения всем тем, кто стремится к воздействию на общество посредством речевого общения, чрезвычайно большого значения фонетических свойств и характеристик высказывания. Это задача не только воспитания культуры речи говорящего, но и задача сохранения минимально «загрязненной» фонетической среды, в которой развивается и существует литературный язык и его звуковые средства.

Литература

ОБЩИЕ ВОПРОСЫ

Бондарко Л. В. 1) Звуковой строй современного русского языка. М., 1977; 2) Фонетическое описание языка и фонологическое описание речи. Л., 1981.

Бондарко Л. В., Вербицкая Л. А., Гордина М. В. Основы общей фонетики. СПб., 1991.

Зиндер Л. Р. Общая фонетика. М., 1979. 312 с.

Златоустова Л. В., Потапова Р. К., Грунин-Донской В. Н. Общая и прикладная фонетика. М., 1986.

Экспериментально-фонетический анализ речи: Проблемы и методы. Вып. 1. Л., 1984; вып. 2. 1989.

АКУСТИКА РЕЧИ И АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ

Бюллетень фонетического фонда русского языка / Под ред. К. Сапока и Л. Бондарко. Вып. 1. Ленинград; Бохум, 1988; вып. 2. 1989; вып. 3. 1990; вып. 4. 1992.

Деркач М. Ф., Гумецкий Р. Я., Гура Б. М., Чабан М. Е. Динамические спектры речевых сигналов. Львов, 1983.

Кейтер Дж. Компьютеры — синтезаторы речи. М., 1985.

Обжелян Н. К., Трунин-Донской В. Н. Речевое общение в системах «человек — ЭВМ». Кишинев, 1985.

Рамишвили Г. С., Чикоидзе Г. Б. Криминалистическое исследование фонограмм речи и идентификация личности говорящего. Тбилиси, 1991.

Распознавание слуховых образов / Под ред. Н. Г. Загоруйко и Г. Я. Волошина. Новосибирск, 1970.

Сапожков М. А. 1) Речевой сигнал в кибернетике и связи. М., 1963; 2) Электроакустика. М., 1978.

Уровни языка в речевой деятельности: К проблеме лингвистического обеспечения автоматического распознавания речи / Под ред. Л. В. Бондарко. Л., 1986.

Фант Г. Акустическая теория речеобразования. М., 1964.

Фланаган Дж. Анализ, синтез и восприятие речи. М., 1968.

Цемель Г. И. Опознавание речевых сигналов. М., 1971.

РЕЧЕОБРАЗОВАНИЕ

Балонов Л. Я., Деглин В. Л. Слух и речь доминантного и недоминантного полушарий. Л., 1976.

Детский голос / Под ред. В. Н. Шацкой. М., 1970.

Логопедия / Под ред. Л. С. Волковой. М., 1989.

Морозов В. П. Биофизические основы вокальной речи. М., 1977.

Речь: Артикуляция и восприятие / Под ред. В. А. Кожевникова и Л. А. Чистович. М.; Л., 1965.

Скалозуб Л. Г. Динамика звукообразования (по данным кинорентгенографирования). Киев, 1979.

ВОСПРИЯТИЕ

Анализ речевых сигналов человеком / Отв. ред. Г. В. Гершуни. М., 1971.

Бару А. В. Слуховые центры и опознание звуковых сигналов. Л., 1978.

Бондарко Л. В., Загоруйко Н. Г., Кожевников В. А., Молчанов А. П., Чистович Л. А. Модель восприятия речи человеком. Новосибирск, 1968.

Восприятие: Механизмы и модели / Пер. с англ.; Под ред. Н. Ю. Алексеенко. М., 1974.

Механизмы речеобразования и восприятия сложных зуков / Отв. ред. Г. В. Гершуни. М.; Л., 1966.

Трауготт Н. Н., Кайданова С. И. Нарушение слуха при сенсорной алалии и афазии. Л., 1975.

Физиология речи: Восприятие речи человеком. Л., 1976.

Физиология сенсорных систем. Л., 1972.

ПРОЧИЕ ПРОБЛЕМЫ

Воронин С. В. Фоносемантические идеи в зарубежном языкознании. Л., 1990.

Журавлев А. П. Фонетическое значение. Л., 1974.

Русское сценическое произношение / Под ред. С. М. Кузьминой. М., 1986.

МОДЕЛИРОВАНИЕ СИНТАКСИСА ЕСТЕСТВЕННОГО ЯЗЫКА

Синтаксический анализ предложения является необходимым этапом автоматической обработки текста или речи при решении различного рода задач в прикладной лингвистике. Целью синтаксического анализа является распознавание синтаксической структуры предложения или отдельных его частей. В процессе синтаксического анализа устанавливается, удовлетворяет ли цепочка словоформ, образующих предложение, структурным условиям, заданным явным образом правилами грамматики языка. При положительном ответе на этот вопрос анализируемое предложение признается правильным с точки зрения данной грамматики, и его структура получает явное описание, или представление, в форме, принятой в рамках данной грамматики. Разработка способов представления синтаксической структуры предложения — одна из основных проблем теории синтаксических структур.

Прежде чем перейти к теории синтаксических структур, остановимся на некоторых задачах автоматической обработки текста, для решения которых необходим синтаксический анализ, т. е. знание синтаксической структуры полного предложения или отдельных его частей, например, именных словосочетаний в составе предложения.

Автоматический синтаксический анализ выполняется во всех достаточно развитых системах машинного (автоматизированного) перевода с одного естественного языка на другой. Синтаксическая структура здесь может использоваться как для получения семантической интерпретации предложения, так и непосредственно — для преобразования синтаксической структуры входного (переводного) предложения одного языка в структуру выходного предложения другого языка в системах, основанных на так называемом грамматическом переводе. При этом используется набор грамматических соответствий переводимого и переводящего языков.

При анализе текста (текста документов и текста запросов) в информационно-поисковых системах иногда применяется не-

полный, или приближенный, синтаксический анализ, при котором распознается структура отдельных частей предложения, в основном — именных словосочетаний, входящих в предложение. Это обусловлено тем, что целью автоматической обработки текстов в таких системах является индексирование текста — распознавание в тексте документа и запроса наименований понятий, или терминологических словосочетаний, реализующихся в тексте именными словосочетаниями или отдельными словами. При таком анализе используется «грамматика», содержащая описание синтаксической структуры только именных словосочетаний.

Синтаксический анализ предложений естественного языка составляет обычно отдельный блок систем понимания текста в задачах искусственного интеллекта, к которым относится, в частности, обеспечение взаимодействия человека с ЭВМ на естественном языке.

Системы автоматического распознавания и понимания речи также включают в себя блоки семантико-синтаксического анализа сегментных цепочек не до конца распознанных высказываний с целью выдвижения дополнительных гипотез и верификации грамматической и семантической правильности высказываний.

Наконец, отметим и «внутренний лингвистический интерес» к теории синтаксических структур. Например, без точного и формального представления синтаксической структуры предложения невозможны статистические исследования синтаксиса естественных языков по широкому спектру синтаксических параметров.

Основными проблемами моделирования синтаксиса в прикладной лингвистике, как было сказано выше, являются представление синтаксических структур естественного языка и разработка оптимальных методов (алгоритмов) синтаксического анализа, ориентированных на то или иное представление синтаксической структуры.

В данном разделе рассматриваются основные способы представления синтаксической структуры предложения и, в качестве иллюстрации, приводится ограниченный фрагмент формальной грамматики составляющих. Алгоритмы автоматического синтаксического анализа строятся на базе той или иной модели формальной грамматики, но здесь эти проблемы, требующие специального обсуждения, не рассматриваются.

В настоящее время широко используются три основных способа представления синтаксической структуры предложения: структуры составляющих (дерево непосредственно составляющих), структуры зависимостей (дерево зависимостей) и структуры, совмещающие свойства структур зависимостей и составляющих (ориентированные структуры составляющих).

СТРУКТУРА СОСТАВЛЯЮЩИХ

В модели составляющих (или в системе составляющих) предложение рассматривается как линейно упорядоченная цепочка словоформ П. Точнее было бы говорить о цепочке минимальных единиц, в качестве которых могут выступать морфы, словоформы или, в некоторых случаях, единицы, большие, чем одна словоформа (например, сложные союзы, сочетания предлогов с существительными и пр.). Множество С отрезков этой линейно упорядоченной цепочки называется системой составляющих на П, если оно удовлетворяет следующим условиям:

1) Само предложение П входит в С в качестве элемента.

2) Все словоформы (отрезки длиной, равной единице) цепочки П также принадлежат С.

3) Любые два отрезка из С либо не пересекаются, либо один из них содержится в другом (Гладкий, 1985, с. 31).

Полное предложение П и отдельные словоформы включаются в систему составляющих С для формальной законченности определения.

Отрезки предложения, образующие множество С, называются составляющими. Само предложение П называется полной составляющей. Составляющие, не равные отдельным словоформам и не являющиеся полными, интерпретируются обычно как словосочетания.

При такой интерпретации система составляющих может рассматриваться как модель синтаксической структуры предложения, в которой в качестве элементарных синтаксических единиц выступают словосочетания, включая отдельные словоформы, а в качестве структурного отношения между словосочетаниями используется отношение включения одного словосочетания в другое. Запрещение «частичного пересечения» словосочетаний (см. п. 3 определения) является вполне естественным требованием для словосочетаний как синтаксически связанных групп в составе предложения.

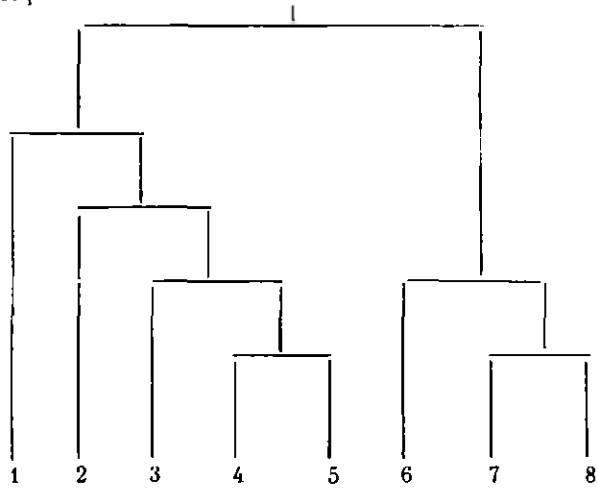
На множестве С можно определить отношение непосредственного вхождения одной составляющей в другую. Выражение «составляющая *B* непосредственно входит в составляющую *A*» означает, что отрезок *B* входит в отрезок *A* и в *C* не существует отрезка *E*, который бы входил в *A* и содержал в себе *B*. Составляющая *B* при этом называется непосредственной составляющей *A*.

Граф, определяемый этим отношением на множестве *C*, является деревом, корнем которого является полная составляющая П, а висячими узлами — словоформы (минимальные единицы). Это дерево называется деревом составляющих, или деревом НС.

Дерево составляющих является одним из наглядных графи-

ческих способов записи системы составляющих С. Дерево составляющих часто называют также структурой составляющих (предложения П).

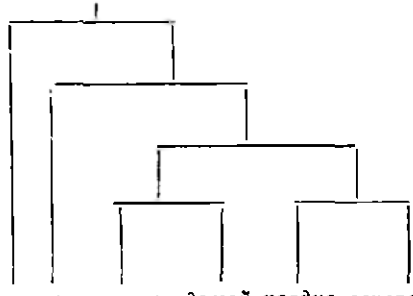
Пример дерева НС предложения *Для представления синтаксической структуры предложения используется структура составляющих:*



Возможна скобочная запись структуры составляющих предложения, при которой каждая неодноэлементная составляющая, включая полную, заключается в скобки:

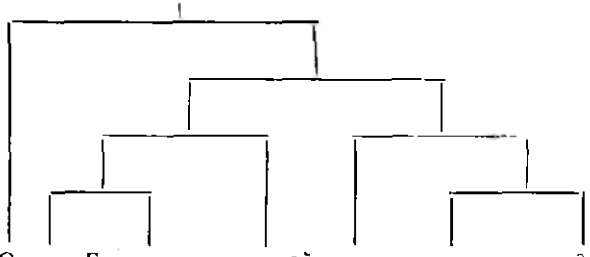
((1 (2 (3 (4 5)))) (6 (7 8)))

В большинстве случаев предложениям русского языка удается приписать вполне «естественные» структуры составляющих, опираясь на «грамматическую» интуицию носителей языка (для русского языка систематического описания грамматики в терминах системы составляющих в настоящее время не существует) и понимание содержания предложения.



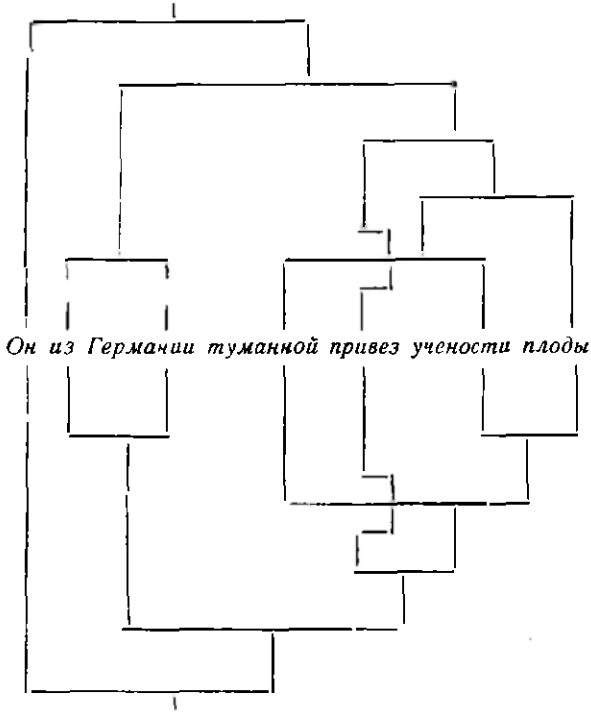
Вчера он вернулся домой поздно вечером

Однако встречаются трудные случаи, когда анализ по составляющим для предложения не очевиден или даже невозможен в пределах точного приведенного выше определения структуры составляющих. На первый взгляд, пушкинская фраза *Он из Германии туманной привез учености плоды* имеет следующую естественную структуру составляющих:

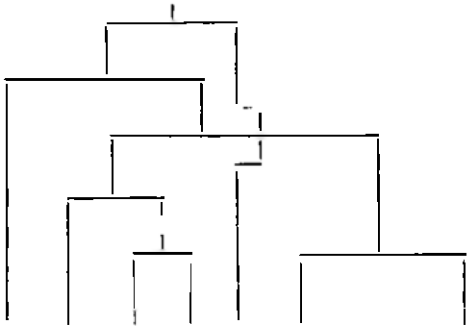


Между тем внимательный анализ содержания фразы заставляет сомневаться в правильности этой структуры как несоответствующей смыслу предложения, который, видимо, точнее выражается высказыванием *Он привез из Германии плоды туманной учености.* Германия никогда не характеризовалась как туманная страна, в отличие от Англии (Гладкий, 1985, с. 25).

Структура составляющих фразы Пушкина, отвечающая данному ее пониманию, может быть двоякой:

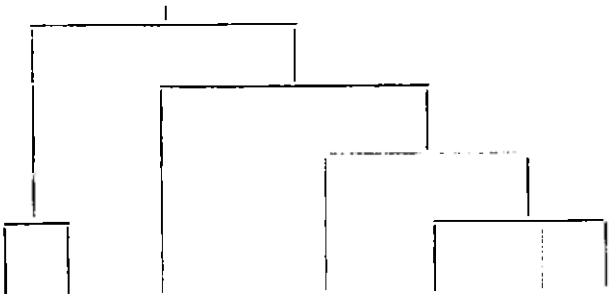


В этой структуре составляющая *туманной... учености плоды* не является отрезком в линейном порядке слов предложения, и вся структура формально не удовлетворяет определению системы составляющих. Такие структуры составляющих, в которых по крайней мере одна из составляющих не является отрезком, называются разрывными структурами составляющих. Разрывные структуры составляющих встречаются в предложениях естественного языка, особенно в текстах художественных, где они часто используются как особый стилистический прием или диктуются ритмической структурой фразы (пример из А. Блока):

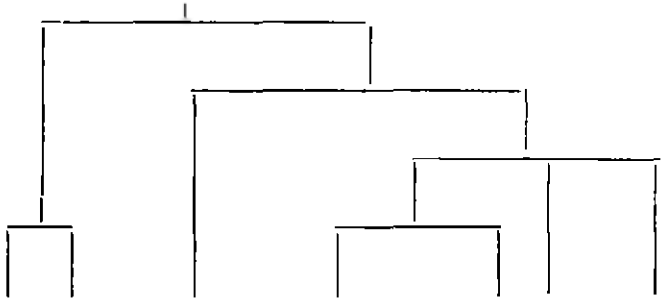


Решал все тот же я мучительный вопрос

Несмотря на трудности анализа по составляющим отдельных предложений, которые связаны в основном с нарушением нормативного порядка слов в предложении, системы составляющих вполне адекватно отражают структуру предложения как с грамматической, так и с семантической точки зрения. Это выражается, например, в том, что семантически неоднозначные (омонимичные) предложения могут быть представлены различными структурами составляющих, соответствующими различным интерпретациям предложения.



1) На вечере присутствовали известные артисты и кинокритики



2) На вечере присутствовали известные артисты и кинокритики

Структура (2) соответствует интерпретации предложения *На вечере присутствовали известные артисты и какие-то кинокритики* (необязательно известные). Структура (1) соответствует значению предложения *На вечере присутствовали как известные артисты, так и известные кинокритики.*

Способность различать на синтаксическом уровне омонимию предложения является одним из существенных достоинств модели систем составляющих.

Лингвистическая значимость модели систем составляющих существенно усиливается, если в структуру составляющих, описывающую структуру предложения, внести дополнительную информацию о типах составляющих, т. е. о типах слов и словосочетаний, или конструкций, входящих в предложение.

Систему составляющих, в которой каждая составляющая имеет метку (имя) класса слова или словосочетания, называют размеченной структурой составляющих.

Так называемые порождающие грамматики составляющих (Гладкий, Мельчук, 1969; Фитиалов, 1984) приписывают порождаемым предложениям именно размеченные деревья составляющих. Грамматики составляющих содержат правила «развертывания» синтаксических конструкций в составляющие нижележащих уровней и, с формальной точки зрения, представляют собой правила подстановки вместо конструкций определенного типа конструкций, входящих в их состав по правилам грамматики данного языка.

В качестве иллюстрации грамматики составляющих приведем фрагмент такой грамматики, порождающей предложения фиксированной структуры: *В это время за окном раздался очень громкий шум.*

«Лексическое наполнение» предложений данной грамматической структуры может быть различным, например: *В последние дни за кулисами шел неоправданно бурный спор.*

Грамматика составляющих моделирует синтаксическую правильность порождаемых предложений, а не их содержание.

Введем следующие грамматические типы составляющих:

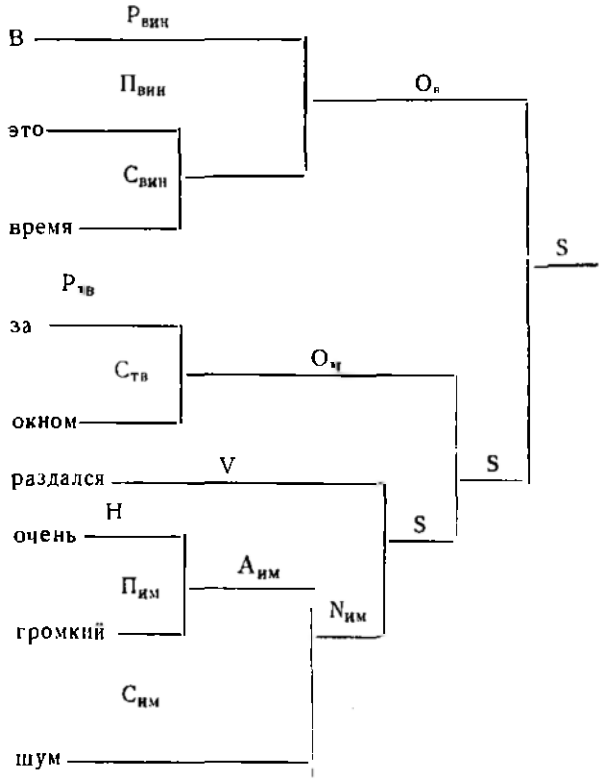
S — предложение,
$O_в$ — обстоятельство времени,
$O_м$ — обстоятельство места,
$N_{вин}$ — группа существительного в вин. падеже,
$N_{им}$ — группа существительного в им. падеже,
$A_{им}$ — группа прилагательного в им. падеже,
$P_{вин}$ — предлог с управлением на вин. падеж,
$P_{тв}$ — предлог с управлением на тв. падеж,
$П_{им}$ — прилагательное в им. падеже,
$П_{вин}$ — прилагательное в вин. падеже,
$C_{им}$ — существительное в им. падеже,
$C_{вин}$ — существительное в вин. падеже,
$C_{тв}$ — существительное в тв. падеже,
V — глагол в личной форме,
H — наречие.

Правила подстановки грамматики составляющих для предложений указанного выше типа:

$S \rightarrow O_вS$
$S \rightarrow O_мS$
$S \rightarrow VN_{им}$
$O_в \rightarrow P_{вин}N_{вин}$
$O_м \rightarrow P_{тв}C_{тв}$
$N_{вин} \rightarrow П_{вин}C_{вин}$
$C_{им} \rightarrow$ шум, спор...
$C_{вин} \rightarrow$ время, дни...
$C_{тв} \rightarrow$ окном, кулисами...
$V \rightarrow$ раздался, шел...
$H \rightarrow$ очень, неоправданно...
$N_{им} \rightarrow A_{им}C_{им}$
$A_{им} \rightarrow НП_{им}$
$P_{вин} \rightarrow$ в...
$P_{тв} \rightarrow$ за...
$П_{им} \rightarrow$ громкий, бурный...
$П_{вин} \rightarrow$ э:о...

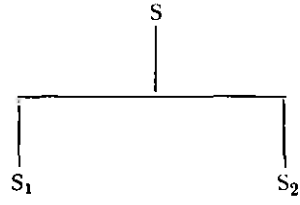
(Фитиалов, 1984, с. 91)

Применение подобных правил грамматики позволяет описывать синтаксическую структуру предложений в виде следующей размеченной системы составляющих:

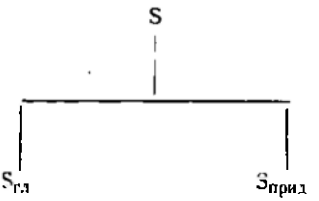


Некоторые правила анализа предложений по составляющим являются в настоящее время общепринятыми. Мы не имеем возможности описать здесь способы анализа по составляющим достаточно широкого набора конструкций грамматики русского языка.

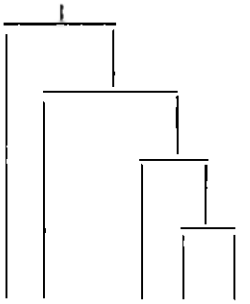
1) Сложносочиненные предложения S_1 и S_2 представляются структурой:



2) Сложноподчиненные предложения объединяются в структуру:

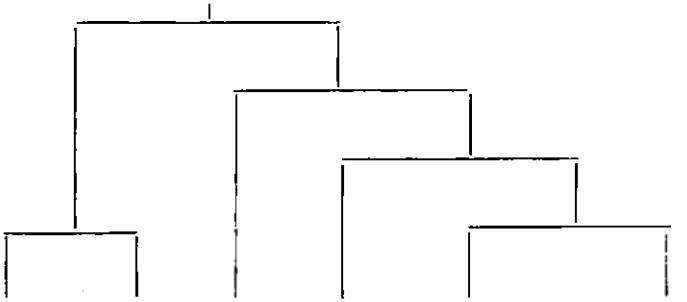


Если придаточное предложение относится не ко всему главному предложению $S_{гл}$, а к отдельному слову-хозяину в составе главного предложения, то оно объединяется с этим словом в отдельную составляющую.

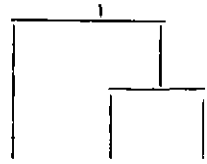


Он знал, что был прав

В словосочетаниях существительного с согласованным и несогласованным определениями вопрос о степени «близости» синтаксической связи между ними обычно решается по семантическим критериям: если согласованное определение (прилагательное, причастие и др.) с существительным обозначает некоторое понятие (устойчивое словосочетание), они прежде всего выделяются в отдельную составляющую, несогласованные определения присоединяются к этой составляющей на следующем уровне. Если понятие (устойчивое словосочетание) образует существительное с несогласованным определением, они объединяются в составляющую, к которой затем присоединяются имеющиеся свободные согласованные определения.



1) цифровые вычислители со встроенным устройством контроля



2) текстовая база данных

СТРУКТУРА ЗАВИСИМОСТЕЙ

На линейно упорядоченной цепочке словоформ предложения П задается бинарное отношение зависимости, отражающее содержательно тот факт, что две словоформы в предложении, которые образуют словосочетание, обычно «неравноправны» и в грамматическом, и в смысловом плане. Это отношение зависимости можно рассматривать как обобщение традиционно выделяемых в предложении отношений управления, согласования и примыкания. Набор таких зависимостей между словоформами предложения дает существенную информацию о синтаксической структуре этого предложения. Эмпирические наблюдения показывают, что в большинстве случаев этот набор зависимостей на множестве вхождений словоформ предложения П может быть представлен в виде графа, являющегося деревом.

Деревом зависимостей предложения П называется конечный граф на множестве вхождений словоформ (узлов) предложения П такой, что:

1) существует единственная словоформа, не зависящая ни от какой другой (эта словоформа называется корнем дерева, или, по лингвистической терминологии, его вершиной),

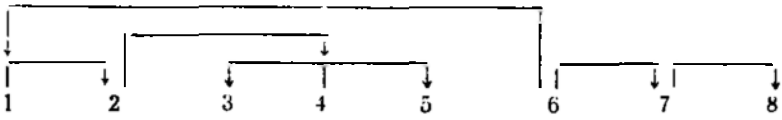
2) всякая словоформа, отличная от стоящей в вершине, зависит ровно от одной другой словоформы,

3) в графе отсутствуют замкнутые пути.

Путь в дереве зависимостей образует последовательность словоформ, в которой каждая следующая словоформа зависит от предыдущей. Для каждого узла дерева существует единственный путь от вершины дерева. Замкнутым путем называется такой путь, в котором начало его и конец совпадают. Если от словоформы x в предложении П ведет путь к словоформе y, то говорят, что словоформа x косвенно подчиняет словоформу y (или: словоформа y косвенно зависит от словоформы x). Употребляют также термин «x руководит словоформой y». Множество всех словоформ, косвенно зависящих от словоформы x, называется группой зависимости этой словоформы. Заметим, что непосредственная зависимость y от словоформы x является частным случаем косвенной зависимости.

Отношение зависимости между словоформами обозначают дугой $X \rightarrow Y$.

Пример зависимостей предложения *Для представления синтаксической структуры предложения используется дерево зависимостей:*



В группу зависимости предлога *для* входит словосочетание *для представления синтаксической структуры предложения.* В группу зависимости вершины дерева *используется* входит все предложение.

Как и структуры составляющих, деревья зависимостей обладают достаточной «различительной силой» для разрешения структурной омонимии предложений.

Структура зависимостей (1) соответствует смыслу предложения: *Он вынул трубку, сделанную из глины,* структура (2): *Он вынул трубку, лежавшую в глине.*

При анализе предложений естественного языка часто используют так называемые размеченные деревья зависимостей. В размеченном дереве зависимостей каждая дуга (отношение зависимости) отнесена к определенному типу грамматической связи между словоформами в предложении. Так, например, отношение между подлежащим и сказуемым — предикативное (пред.), между существительным и согласованным с ним прилагательным — определительное (опред.), между существительным и несогласованным определением — атрибутивное (атриб.), между глаголом и связанным с ним обстоятельством — обстоятельственное (обст.) и др. Набор типов грамматических отношений зависит от конкретной грамматики языка. В своей работе И. А. Мельчук (1964) рассматривает 31 тип зависимостей в русском языке (он называет их отношениями непосредственной доминации).

Размеченное дерево зависимостей для приведенного выше предложения будет таким:

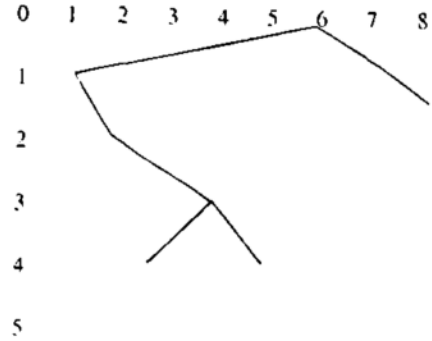
Размеченные структуры зависимостей позволяют различать омонимию предложений или конструкций в их составе в том случае, когда не удается приписать омонимичному предложению два различных неразмеченных дерева зависимостей.

(«Его прислал к нам инспектор»)

(«Его прислали к нам в качестве инспектора»)

Существует еще одно графическое представление структуры зависимостей в виде так называемого расположенного дерева зависимостей (Севбо, 1981). Этот способ изображения структуры предложения состоит в том, что выбирается двумерная система координат, по одной из осей которой (оси абсцисс) располагаются словоформы в линейном порядке их в предложении, а по другой оси (оси ординат) размещаются узлы дерева зависимостей в соответствии с уровнем, занимаемым данным узлом в дереве. При этом ось ординат обычно направлена сверху вниз, так что вершина (корень) дерева зависимостей по лингвистической терминологии остается вершиной дерева (дерево «растет» сверху вниз).

Структура предложения *Для представления синтаксической структуры предложения используется дерево зависимостей* в виде расположенного дерева зависимостей в двумерной системе координат (ср. одномерное изображение «стрелочной» структуры этого предложения, приведенное выше) будет выглядеть таким образом:



Поскольку иерархия узлов структуры зависимостей определяется осью ординат, стрелки при изображении расположенного дерева зависимостей можно не использовать.

ПРОЕКТИВНЫЕ СТРУКТУРЫ ЗАВИСИМОСТЕЙ

Формальное представление структуры предложения в виде дерева зависимостей позволяет строго сформулировать одну из существенных синтаксических закономерностей естественных языков. Речь идет о так называемом свойстве проективности структур зависимостей предложений естественных языков.

Выше было определено понятие группы зависимости словоформы x в предложении П. Напомним, что в группу зависимости словоформы x входят все словоформы предложения П, зависящие непосредственно или косвенно от x.

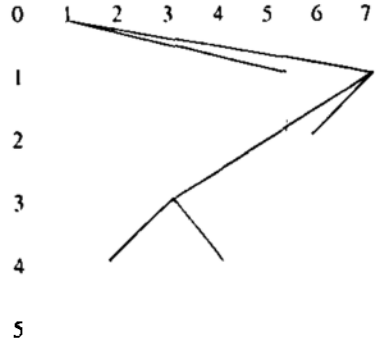
Проективной структурой зависимостей предложения П называется такая структура, в которой группа зависимости каждой словоформы является неразрывным отрезком в линейном порядке слов предложения.

Это свойство структуры предложения известно также как свойство «компактности», когда синтаксически связанные словоформы предложения (словосочетания) расположены в нем связно, образуют неразрывный отрезок предложения.

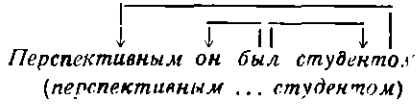
В расположенном дереве зависимостей нарушение свойства проективности формально выражается в том, что линии проекций, ведущие от словоформ, линейно расположенных по оси абсцисс, пересекают дуги дерева зависимостей, отражающие структурные отношения между словоформами в составе предложения, или сами дуги расположенного дерева зависимостей пересекаются между собой (см. приведенную ниже схему).

При одномерном (линейном) представлении дерева зависимостей, когда одновременно учитываются как линейный порядок слов в предложении, так и структурные отношения между словоформами, формальным признаком нарушения законо-

мерности проективности является пересечение стрелок (дуг) зависимостей между словоформами.

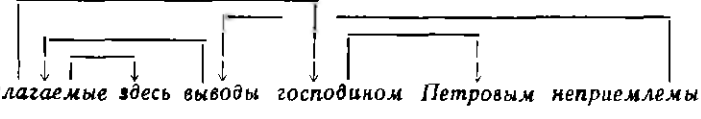


Частным случаем нарушения проективности является «накрытие» стрелкой вершины предложения. При этом не наблюдается пересечения дуг дерева зависимостей, но нарушается основной признак проективности: в предложении появляются «разрывные» группы зависимости отдельных словоформ (т. е. группы, не являющиеся отрезками в линейном порядке слов предложения).



Свойство проективности структуры зависимостей является статистически подтвержденной сильной закономерностью естественного языка, и нарушается оно обычно в двух случаях:

а) в предложениях с явным нарушением нормированного порядка слов («в грамматически неправильных предложениях»):



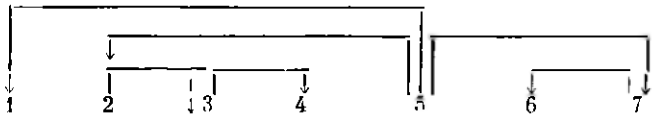
б) в предложениях, когда проективность нарушается с целью достичь определенного художественного (стилистического) эффекта:

И когда в этой хижине тесной
Я, бедняк обездоленный, жду,
Повторяя
напев
неизвестный
в
соловьином,
звенящем
саду

(А. Блок. Соловьиный сад)

Проективность как закономерность синтаксиса естественного языка, которая регулирует отношения линейного порядка и структурные связи между словоформами в предложении, оказывается иногда настолько сильной, что носители языка предпочитают проективную синтаксическую структуру непроективной в ущерб правильной смысловой интерпретации предложения.

Рассмотренное выше предложение *Он из Германии туманной привез учености плоды* (1 2 3 4 5 6 7) интерпретируется обычно как проективное:



Но структура, соответствующая правильному пониманию данного предложения, будет другой, к тому же она будет непроективной:

(Гладкий, 1985)

(Ср. анализ данного предложения по составляющим.)

СОГЛАСОВАНИЕ СТРУКТУР СОСТАВЛЯЮЩИХ И ЗАВИСИМОСТЕЙ

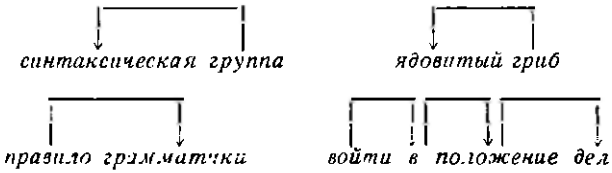
Структуры составляющих и зависимостей в принципе отражают существенно различную информацию о структуре предложения.

Основные различия этих двух представлений синтаксической структуры предложения состоят в следующем.

1) Элементарными единицами в дереве зависимостей являются вхождения словоформ (допускается объединение в один узел дерева сложных предлогов и др.), элементарными синтаксическими единицами в структуре составляющих являются словосочетания (включая в качестве частного случая отдельные словоформы и полное предложение).

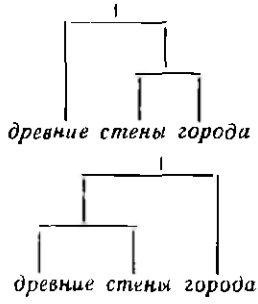
2) Структура зависимостей базируется на «неравноправном» отношении между двумя словоформами x и y, когда одна из них зависит от другой (отношение синтаксической связи является ориентированным, $x \rightarrow y$ или $y \rightarrow x$). В структуре составляющих отношения между непосредственно составляющими некоторой составляющей A не иерархизованы, т. е. среди них не выделяется ни одна непосредственно составляющая в качестве главной (основной), от которой зависели бы другие непосредственно составляющие. Это вполне логично для словосочетаний, в которых выделение главных элементов не требуется — примером таких словосочетаний являются однородные группы — именные или глагольные:

Но в то же время многие синтаксические отношения между элементарными единицами предложения требуют выделения главной и подчиненной единицы:



3) Структура составляющих допускает возможность объединять в одну составляющую «семантически близкие» словосочетания, когда структура зависимостей не может различить подобные смысловые оттенки. Примером здесь может служить анализ словосочетания *древние стены города,* для которого возможна только одна структура зависимостей:

но две возможные структуры составляющих:



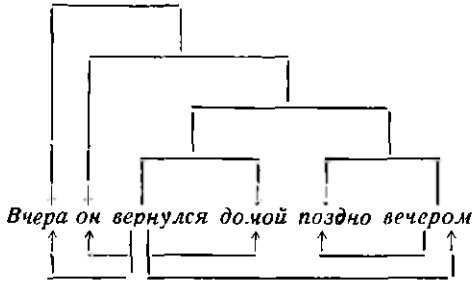
Вопрос о том, насколько существенно различать подобные смысловые оттенки при анализе предложения, мы здесь обсуждать не будем.

Тем не менее, возможность определить существенную связь между двумя моделями представления синтаксической структуры предложения — в виде структуры составляющих и в виде структуры зависимостей — имеется. Для этого необходимо ввести в структуру составляющих дополнительную информацию об ориентации синтаксических связей, а именно — в каждой составляющей *A* выделить среди ее непосредственно составляющих одну в качестве главной непосредственно составляющей, считая остальные составляющие в составе *A* неглавными.

Эта процедура приписывания информации для *A* о главной непосредственно составляющей может выполняться формально, но для конкретной грамматики языка можно пользоваться некоторыми содержательными критериями.

Система составляющих, в которой для каждой составляющей среди ее непосредственно составляющих выделена одна в качестве главной, называется ориентированной структурой составляющих.

Графическое изображение ориентированной структуры составляющих может быть таким — стрелки отмечают неглавные составляющие. Очевиден переход от ориентированной структуры составляющих к структуре зависимостей.



Структура составляющих предложения П и структура зави-

симостей этого предложения называются согласованными, если:

1) группы зависимости всех узлов структуры зависимости являются составляющими структуры предложения П;

2) каждая составляющая структуры предложения П является группой зависимости или усеченной группой зависимости некоторого узла структуры зависимости предложения П.

Под усеченной группой зависимости узла *x* в дереве зависимости предложения П понимается произвольная часть полной группы зависимости этого узла, например, *вернулся вчера, вернулся домой поздно вечером* и др. При включении усеченных групп в структуру составляющих предложения П необходимо соблюдать правило (3) в определении системы составляющих, а именно: усеченные группы в качестве составляющих не должны частично пересекаться (допускается только включение одной усеченной группы в другую или их непересечение).

Литература

Гладкий А. В. 1) Синтаксические структуры естественного языка в автоматизированных системах общения. М., 1985; 2) Формальные грамматики и языки. М., 1973.

Гладкий А. В., Мельчук И. А. Элементы математической лингвистики. М., 1969.

Мельчук И. А. Автоматический синтаксический анализ. Новосибирск, 1964.

Падучева Е. В. О способах представления синтаксической структуры предложения // Вопр. языкозн. 1964. № 2.

Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика. Киев, 1981.

Фитиалов С. Я. Формальные грамматики. Л., 1984.

Шрейдер Ю. А. Равенство, сходство, порядок. М., 1971.

МОДЕЛИРОВАНИЕ СЕМАНТИКИ ПРЕДЛОЖЕНИЯ

1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Семантика предложения стала объектом интенсивных исследований в 60-е годы нынешнего века. Спорадически отдельные ученые обращались к этой теме и раньше, однако преимущественно как к попутной, в связи с обсуждением предложения и его типов в синтаксисе. Как самостоятельная область исследования синтаксическая семантика три десятилетия назад еще не существовала.

Развитию семантики предложения способствовало несколько благоприятных обстоятельств, среди которых можно отметить три наиболее важных. Одно из этих обстоятельств обусловлено внутренней логикой развития самой лингвистики как науки, второе можно рассматривать как внешний стимул, а третье — как результат взаимодействия лингвистики со смежными областями знания.

Влияние первого обстоятельства проявилось в осознании многими лингвистами той простой истины, что анализ формальной структуры предложения, как и языка вообще, сколь бы совершенным и тонким он ни был, не помогает в достаточной степени проникнуть в природу языковой коммуникации человека, которая представляет собой прежде всего обмен содержательной информацией, смыслом. Следовательно, изучению необходимо подвергнуть сам этот смысл. Уровень лингвистической науки середины XX в. оказался достаточным для того, чтобы она могла заняться решением данной проблемы.

Роль внешнего стимула для развития синтаксической семантики сыграла та сфера исследований, которая получила в дальнейшем название искусственного интеллекта. Здесь имеются в виду идеи о сходстве ряда мыслительных действий человека и компьютера, восходящие к Н. Винеру, А. Тьюрингу, К. Штейнбуху, Ф. Джорджу и др., а также сформулированный несколько позже тезис о том, что интеллектуальный автомат

должен быть в состоянии вести общение с человеком на естественном языке либо в форме устной речи, либо в текстовой форме. Известно, что такие автоматы были в дальнейшем построены.

Что касается взаимодействия лингвистики со смежными отраслями знания, то в наиболее отчетливой форме оно проявилось в отношении таких наук, как философия, социология, психология, информатика и др. Однако самое плодотворное влияние на лингвистику и особенно на синтаксическую семантику оказала символическая логика. Идеи и терминология современной логики широко используются в настоящее время в исследованиях по синтаксической семантике.

Говоря о синтаксической семантике, не следует забывать и обычные работы по синтаксису. Дело в том, что многие синтаксисты сделали большой шаг вперед не только в области исследования формальных синтаксических структур, но и в сфере содержательного синтаксиса. Роль важного пролога к интенсивной разработке проблемы семантики предложения сыграли труды Л. Теньера, Е. Куриловича, Ш. Балли, а также некоторые идеи трансформационной порождающей грамматики. Особенно большое значение имела вербоцентрическая концепция предложения Л. Теньера, в которой конструктивная целостность предложения связывалась с валентностью глагола-сказуемого. Принцип вершинности, доминантности глагола, подчиняющего себе актанты, хорошо коррелировал с принципом функциональной значимости предиката в логике, который в рамках пропозиции выступает как функция от некоторого числа аргументов, или переменных.

При моделировании семантики предложения прежде всего возникает вопрос о том, что считать смыслом предложения. На этот вопрос однозначного ответа в настоящее время не существует. Однако большинство исследователей сходится на том, что смысл предложения представляет собой сложное, многоаспектное образование. В содержании предложения сложнейшим образом сфокусированы характеристики экстралингвистической действительности, ее отражения в сознании человека в виде концептуальных структур, коммуникативных установок участников общения, а также особенности самого языка. При моделировании семантики предложения каждый из названных аспектов может получить статус исходного или даже единственного, если прочие аспекты по каким-либо соображениям не рассматриваются. В существующих теориях синтаксической семантики так дело и обстоит. Тем самым в центре рассмотрения оказывается какой-нибудь один аспект, в то время как прочим аспектам отводится второстепенная роль. Впрочем, изредка на равных правах учитывается два аспекта.

В зависимости от того, какой исходный аспект кладется в основу моделирования семантики предложения, возможны в

принципе четыре подхода. Первые три подхода учитывают элементы треугольника соотнесенности «мир — мышление — язык», четвертый основывается на прагматическом отношении «язык — человек». Эти подходы будут в дальнейшем именоваться следующими терминами: 1) онтологоцентрический, 2) концептоцентрический, 3) синтактикоцентрический, 4) антропоцентрический. В рамках каждого из перечисленных здесь подходов существует некоторая система воззрений, разработанных с различной степенью детальности. Поэтому существенно выявить фундаментальные принципы и постулаты, которыми можно руководствоваться, строя теорию с преимущественной ориентацией на какой-нибудь один аспект.

2. ОНТОЛОГОЦЕНТРИЧЕСКИЙ ПОДХОД

2.1. ПОНЯТИЕ СИТУАЦИИ

Сущность онтологоцентрического подхода заключается в том, что исходным пунктом для построения теории смысла предложения считается онтология, т. е. отражаемая языком экстралингвистическая действительность. Минимальным компонентом ее, который может описываться предложением, принято считать ситуацию, или факт. В соответствии с этим исследование смысла предложения обычно начинается с определения понятия ситуации и с классификации ситуаций, каждой из которых ставится в соответствие определенная семантическая структура, лежащая в основе того или иного предложения.

Определения ситуации, которые встречаются в научной литературе, не отличаются достаточной ясностью. Ситуация, по мнению Л. Витгенштейна, есть «то, что имеет место» (Витгенштейн, 1958). И. П. Сусов называет ситуацией «ансамбль взаимосвязанных элементов» (Сусов, 1973, с. 15—16), в качестве которых выступают предметы и признаки. По определению В. Г. Гака, «ситуация есть отрезок, часть отражаемой в языке действительности, т. е. движущейся материи» (Гак, 1973, с. 359). Она включает в себя материальные объекты, или субстанции, и формы их координации, т. е. пространство и время.

Моделирование семантики предложения при движении от ситуации к предложению производится следующим образом. В реальной действительности признается существование единичных, или эмпирических, ситуаций. Эти ситуации могут различаться по структуре, а именно по характеру признака или координации, а также по числу и характеру материальных объектов. Индивидуальная эмпирическая ситуация описывается с помощью конкретного высказывания. Вследствие этого высказывание должно содержать компоненты, соответствующие компонентам ситуации. Кроме того, в высказывании должны найти отражение и все отношения, существующие между компонентами ситуации.

Конкретные эмпирические ситуации обобщаются сознанием и сводятся в классы, или в абстрактные ситуации. Абстрактной ситуации соответствует не высказывание, а предложение. Эта ситуация в принципе структурируется так же, как и конкретная. Ситуация простейшего типа называется элементарной ситуацией, элементарным событием или атомарным фактом. На языковом уровне такая ситуация описывается элементарным, или ядерным, предложением. Отображением структуры элементарной ситуации как онтологической сущности на элементарное предложение как синтаксическую сущность является концептуальное образование, именуемое разными авторами терминами «реляционная структура», «глубинная структура», «предикатное выражение» или «атомарная пропозиция».

Разумеется, ситуация может иметь и более сложное строение, т. е. она может складываться из нескольких элементарных ситуаций. В этом случае ей будет соответствовать более сложная пропозициональная структура, лежащая в основе более сложного синтаксического образования.

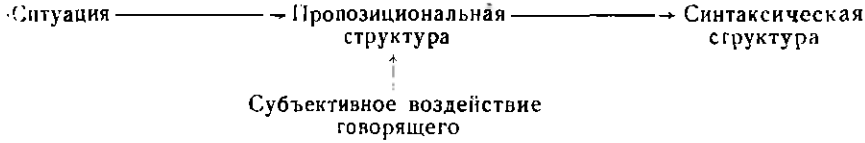
Таким образом, при онтологоцентрическом моделировании порядок движения от мира к языку можно представить в виде следующих трех основных ступеней:

Ситуация (Онтология) —1→ Пропозициональная структура (Смысл) —2→ Синтаксическая структура (Языковое выражение)

Эти ступени соответствуют сферам реальной действительности, мышления и языка, причем особый интерес в них представляют переходы *1* и *2*. Переход *1* связан с отражением ситуации в сознании. Ситуация при этом схематизируется, однако пропозициональная структура изоморфна структуре ситуации. Эта изоморфность обеспечивается тем, что каждому компоненту ситуации соответствует определенный элемент пропозициональной структуры. Действительно, в рамках элементарной пропозиции выделяется предикат, являющийся концептуальным аналогом онтологического признака (действия, процесса, свойства, отношения), и аргументы, именуемые иногда реальными, или семантическими, актантами. Аргументы соответствуют объектам, или так называемым партиципантам ситуации. Поскольку каждый партиципант в рамках ситуации выполняет какую-то роль — активную или пассивную, эта ролевая характеристика находит свое отражение и в пропозициональной структуре в форме семантических ролей типа агенса, пациенса, бенефицианта, локатива и т. д. В пропозициональной структуре эти роли указывают на смысловое отношение аргумента к предикату.

В случае не элементарной, а более сложной ситуации пропозициональная структура складывается из нескольких взаимодействующих друг с другом пропозиций.

В отличие от перехода *1* переход *2* не изоморфен и представляет собой несравненно более сложный процесс, поскольку на этом этапе в действие вступает субъективный фактор говорящего, проявляющийся в конкретной фокусировке ситуации, в соответствии с которой определенный аргумент выбирается в позицию субъекта. Кроме того, на пропозициональную структуру говорящий накладывает содержание своего коммуникативного намерения, собственную модальность, собственную оценку важности компонентов пропозиционального содержания, а также содержательные элементы, отражающие имеющуюся у говорящего модель слушающего. Только с учетом всего этого возможна реализация пропозициональной структуры в форме синтаксического объекта, т. е. предложения, ибо характер предложения не может не зависеть от названных здесь субъективных факторов, не имеющих прямого отношения к самой ситуации как таковой. Таким образом, приведенная выше схема усложнится и будет выглядеть следующим образом:



Естественно, что преобразующее влияние говорящего на пропозицию исключает полную изоморфность пропозициональной и синтаксической структур. Поэтому переход *2* оценивают обычно отношением изосемии, или семантической изоморфности предложения и пропозиции. Изосемия выражается в том, что среди множества синтаксических структур, соответствующих одной и той же пропозиции, имеется одна структура, которая в минимальной степени маркирована субъективным воздействием говорящего и, таким образом, в наиболее чистом виде отражает пропозицию. Такую синтаксическую структуру называют обычно ядерным, или примарным, предложением (Адамец, 1978). Все прочие предложения рассматриваются как трансформы ядерного. Ядерное предложение и его предложенческие трансформы представляют собой разновидности элементарного предложения. Возможны, разумеется, и непредложенческие трансформы ядерного предложения.

2.2. ПОНЯТИЕ РЕФЕРЕНЦИИ

В рамках онтологоцентрического подхода можно выделить концепции, в которых структура экстралингвистических ситуаций не анализируется, но которые учитывают референцию, т. е. связь языковых выражений с действительностью, и языковые механизмы, обеспечивающие такую связь (Падучева, 1985).

Считается, что референция возможна только на уровне конкретных высказываний, т. е. в речи, в то время как механизмы референции являются принадлежностью языка. Референциальными свойствами могут обладать пропозиции и именные группы, занимающие в пропозициях место аргументов. Предназначенность пропозиции или именной группы к тому или иному типу соотнесения с реальной действительностью рассматривается как их денотативный статус. Например, пропозиции могут иметь утвердительный, пресуппозиционный и нейтральный статусы, а именные группы могут иметь нереферентное и референтное употребление, причем в последнем случае референция может быть определенной или неопределенной.

Поскольку референция возможна только в процессе коммуникативного акта, т. е. в речи, то она определенным образом связана с прагматикой. Соответственно этому содержание высказывания должно включать в себя не только семантический компонент, но и прагматический. Первая часть, или первый компонент, представлена пропозицией, которая рассматривается как «потенциальный концепт предложения». Второй компонент складывается из прагматических переменных, к числу которых относятся говорящий, слушающий, место и время речевого акта, общее поле зрения и общий фонд знаний. При замене переменных константами потенциальный концепт предложения превращается в замкнутую пропозицию, способную иметь истинностное значение. Сходные проблемы анализируются также в референциальной логике.

Рассмотренную здесь модель целесообразно относить к онтологоцентрическому подходу, хотя здесь исследуется не онтология, а только связь онтологии с языком. Дело в том, что любой выход языка в экстралингвистический мир является основой онтологоцентрической концепции. Все онтологоцентрические концепции имеют неоспоримое достоинство, поскольку они исходят из важного методологического принципа, гласящего, что язык есть средство отражения и описания окружающего мира. Изучение языка, в том числе и его семантики, вне связи с реальной действительностью рассматривается в настоящее время как неприемлемое. В онтологоцентрических концепциях в определенной степени учитывается и субъективная переработка пропозиции под действием требований коммуникации, что находит свое выражение в выборе изосемичных и неизосемичных конструкций, а также в учете прагматических переменных.

Однако у онтологоцентрического направления есть некоторые уязвимые места. Первый и самый главный его недостаток заключается в том, что экстралингвистические ситуации систематизируются и классифицируются в соответствии с выражающими их типами ядерных предложений. Остается неясным, можно ли вычленять и упорядочивать ситуации независимо от

языка. Имеются и другие недостатки. Дело в том, что не всякое предложение отражает положение дел в окружающем мире. Например, существуют предложения и целые тексты (сказки, мифы, легенды и пр.), которые не имеют каких-либо коррелятов в реальной действительности, а онтологический статус некоторых предложений, которые, по идее, должны описывать положение дел в мире (например, отрицательных, ср. Бондаренко, 1983), не вполне ясен. Именно поэтому некоторые исследователи отказываются от рассмотрения онтологического уровня и от анализа ситуаций и начинают изучение содержательной стороны предложения с концептуального уровня.

3. КОНЦЕПТОЦЕНТРИЧЕСКИЙ ПОДХОД

3.1. РАЗНОВИДНОСТИ КОНЦЕПТОЦЕНТРИЧЕСКОГО ПОДХОДА

В рамках концептоцентрического подхода вопрос об организации экстралингвистического мира не ставится. Исходным пунктом рассмотрения в теориях этого типа является не ситуация, а концептуальная структура, лежащая в основе предложения. Указанная структура извлекается не из экстралингвистических фактов, а непосредственно из данных языка.

Этот подход представлен большим количеством теорий, которые могут быть объединены в несколько направлений. Специфика каждого направления будет определяться используемой логикой и некоторыми другими дополнительными факторами.

Самыми распространенными из концептоцентрических теорий в настоящее время являются теории, основывающиеся на некоторых понятиях символической логики. При этом ключевым понятием оказывается понятие пропозиции. В логике под пропозицией понимают концептуальную структуру, построенную по модели функции от одной или нескольких переменных, которые в рамках пропозиции должны быть либо связаны (на переменные должны быть навешены кванторы), либо специфицированы (переменные должны быть заменены константами). В предложениях естественного языка функции, кванторы и константы выражаются соответствующей лексикой или грамматикой.

В лингвистике пропозиция определяется как «предикат в сочетании с необходимой для выявления его семантики актантной рамкой» (Арутюнова 1976а, с. 52) или как «конфигурация лексических наименований, взаимосвязанных семантико-синтаксическими отношениями в одно смысловое целое» (Адамец, 1978, с. 7). Кроме того, лингвисты привносят в понятие пропозиции и многие другие элементы, считающиеся избыточными при логическом анализе пропозиции

Все теории, в которых при моделировании семантики предложения используется понятие пропозиции, будут условно относиться к пропозициональному направлению. В рамках этого направления можно выделить теории, в которых для характеристики переменных (аргументов) не используются падежно-ролевые отношения, а предикаты могут либо подвергаться, либо не подвергаться семантической декомпозиции. Имеются также теории, опирающиеся на принцип падежно-ролевых отношений.

Моделирование семантики предложения может производиться не только на основе пропозиционального принципа, но и на базе модели суждения, как она интерпретируется в общей, или аристотелевской, логике. Суждение в этой логике определяется как мыслительная структура, состоящая из субъекта, предиката и связки. При этом могут учитываться дополнительные смысловые отношения между субъектом и предикатом, а также другие признаки, которые могут оказаться существенными. Теории, использующие аристотелевскую модель суждения, будут отнесены к субъектно-предикатному направлению.

Выделенные здесь направления требуют более подробного рассмотрения.

3.2. ПРОПОЗИЦИОНАЛЬНОЕ НАПРАВЛЕНИЕ

3.2.1. Теории, не учитывающие падежно-ролевых отношений

К числу теорий, в которых реализован этот принцип, можно отнести теории Е. В. Падучевой (1974), П. Адамца (1978), генеративную семантику (Макколи, 1981; Лакофф, 1981) и некоторые другие.

Например, в концепции П. Адамца пропозиция складывается из предиката и партиципантов, в качестве которых могут выступать компоненты с субстанциальным значением, называемые актантами, и компоненты с обстоятельственным значением, именуемые адвербиальными комплементами. Указанные компоненты представлены в пропозиции основами слов. При этом если в пропозиции имеется несколько актантов или комплементов, то они различаются посредством нумерации. Правила нумерации не сформулированы с необходимой точностью, однако чаще всего на синтаксическом уровне первый актант выражается подлежащим, второй актант — прямым дополнением, третий актант — косвенным дополнением. Переход от пропозиции к предложению осуществляется с помощью операций спецификации и реализации. Средства спецификации называются спецификаторами. Выделяются спецификаторы четырех типов, а именно: модальности, грамматической презентации, актуального членения, а также вида, фазисности и времени. Разграни-

чиваются примарные спецификаторы, дающие примарные, или ядерные, предложения, и непримарные спецификаторы, которые приводят к получению трансформов. Таким образом, спецификация и реализация, под которой понимается вывод на фонологический уровень специфицированной пропозиции, непосредственно самой пропозиции не принадлежат, а относятся к ее грамматическому оформлению. Поскольку грамматическое оформление пропозиции оказывается несравненно более сложным, нежели аппарат построения пропозиций, то основная проблематика, рассматриваемая П. Адамцем, оказывается синтаксической.

В названной выше работе Е. В. Падучевой выделяется уровень *глубинных структур, который фактически* представляет собой уровень пропозиций. Это вытекает из того, что для построения глубинных структур используется язык исчисления предикатов и что «глубинная структура, — по определению Е. В. Падучевой, — должна быть инвариантом относительно системы синтаксических трансформаций» (Падучева, 1974, с. 39). В отличие от теории П. Адамца в работе Е. В. Падучевой более детально и четко сформулированы ограничения на лексику и синтаксис глубинных структур. При характеристике предиката учитываются его семантические валентности, аргументы различаются нумерацией. Глубинные структуры репрезентируются в форме древовидных схем и скобочной записи. П. Адамец пользуется только скобочным представлением.

В генеративной семантике используется два уровня представления пропозиции. Единицы первого уровня называются семантическими представлениями. Специфика их заключается в том, что предикат на этом уровне подвергается семантической декомпозиции, т. е. разложению на семантические признаки, причем каждый такой признак рассматривается как элементарный предикат. Например, предикат «убивать» разлагается на признаки «каузировать+становиться+не+живой». Признаки в семантическом представлении, имеющем вид древовидной схемы, расположены иерархически. Каждый признак имеет собственные аргументы, образуя нечто вроде субпропозиции. При этом каждая субпропозиция, расположенная в иерархии ниже, включена в аргументную позицию верхней смежной субпропозиции.

Единицей второго уровня представления пропозиции является семантическая структура, по отношению к которой осуществлены все трансформации подъема предиката и в которой совокупность семантических признаков заменена лексемой. Аргументы в такой пропозиции различаются нумерацией.

Между первым и вторым уровнями возможны и промежуточные уровни, на которых используются не все допустимые трансформации подъема предиката, в результате чего предикаты на этих уровнях частично состоят из лексем, а частич-

но — из семантических признаков, т. е. имеют вид «каузировать+становиться+мертвый», «каузировать+умирать».

Семантическое моделирование предложения в терминах пропозиций с нумерованными аргументами не получило столь широкого распространения, как моделирование на пропозициональной основе, но с использованием падежно-ролевых отношений. Это связано с тем, что модели второго типа более содержательны, чем первые. Имеются и другие причины, которые будут рассмотрены ниже.

3.2.2. Падежно-ролевые теории

Теории данного типа отличаются исключительным разнообразием. Однако во всех них в качестве основного используется понятие падежной роли. Популярность этого понятия объясняется многими факторами, важнейшими из которых являются следующие.

1. Нумерация аргументов дает возможность достичь только их различения и упорядочения, но не позволяет обеспечить их отождествление, поскольку номерам не приписывается определенная семантика, а это значит, что при одном предикате аргумент с некоторым номером может иметь одно смысловое содержание, тогда как при другом предикате аргумент с тем же номером может иметь совершенно иной смысл. С другой стороны, аргумент, помеченный падежной ролью, имеет вполне определенное смысловое содержание, поэтому пропозиция, аргументы которой снабжены падежными маркерами, содержит больше смысловой информации о предложении, нежели пропозиция с нумерованными аргументами.

2. Использование аппарата падежно-ролевых отношений позволяет более однозначно и экономно описывать семантическую деривацию предложений, так как при любых деривационных преобразованиях предложения характер смыслового содержания аргумента остается неизменным. Нумерация аргументов создает большие затруднения при учете и описании деривационно-семантических отношений.

3. При описании семантики предложения в терминах падежных ролей сохраняется связь с традиционной теорией частных значений падежа, которой занимались такие выдающиеся лингвисты, как Г. Суит, А. де Гроот, Э. Бенвенист, Л. Ельмслев, Р. Якобсон, Е. Курилович и др. В работах этих ученых было сделано много полезных наблюдений относительно того явления, которое в дальнейшем стало именоваться падежной ролью. А это свидетельствует о том, что падежно-ролевое направление в синтаксической семантике имеет давнюю традицию.

Фактически же эта традиция уходит к значительно более далекому прошлому и имеет свои корни в глубокой древности.

Хотя основоположником падежно-ролевого подхода принято считать Ч. Филлмора (1981), идея описания структуры предложения в терминах падежно-ролевых отношений была высказана еще в V в. до н. э. древнеиндийским грамматистом Панини в его знаменитом труде «Восьмикнижие». Панини пользовался понятием kāraka, весьма близкого к понятию глубинного падежа (падежной роли) у Ч. Филлмора. Среди kāraka у Панини фигурировали: kartā — агенс, karma — пациенс или объект, karaṇa — инструмент, adhikaraṇa — место и др. Аналогичные падежные отношения содержатся и в перечне Ч. Филлмора. Поверхностные падежные формы, являющиеся средством выражения падежных ролей, именуются у Панини термином vibhakti (Катенина, Рудой, 1980, с. 78—79). У Филлмора они называются падежными формами. Важно при этом отметить, что описание предложения в терминах kāraka не преследовало у Панини цели описания его смысла. Оно рассматривалось в его работе как обычное синтаксическое описание.

В настоящее время среди исследователей, развивающих падежно-ролевой подход, можно помимо Ч. Филлмора назвать У. Чейфа, Д. Нильсена, У. Кука, Ф. Блейка, С. Старосту, Дж. Андерсона, Р. Шенка, Р. Ван-Валина и У. Фоли, Р. Зимека, Ю. Д. Апресяна, Т. Б. Алисову, И. П. Сусова, В. Г. Гака, В. В. Богданова и др. В качестве эквивалента термина «падежная роль» у названных авторов фигурируют термины «семантическая роль», «семантическая валентность», «семантическая функция аргумента», «референциальная роль», «глубинный падеж», «реальный актант», «семантический актант», «партиципант» и др.

3.2.3. Принципы моделирования семантики предложения при падежно-ролевом подходе

Теории, кладущие в основу описания смысла предложения падежно-ролевой подход, в очень значительной степени отличаются друг от друга как в деталях, так и в некоторых более существенных моментах и в используемой терминологии. Однако во всех существующих теориях можно выделить некоторые очень важные проблемы, требующие решения или частично нашедшие его. Эти проблемы — следующие.

Пропозиция и модус. В падежно-ролевых концепциях принято разграничивать пропозицию и модус. Это разграничение восходит к Ш. Балли, выделявшему диктум (т. е., фактически, пропозицию) и модус. Пропозиция при этом рассматривается как предикатно-аргументная структура, *либо соотносимая с* соответствующей экстралингвистической ситуацией (при онтологоцентрическом подходе), либо постулируемая автономно (при концептоцентрическом подходе). Что касается модуса,

то он обычно выносится за пределы пропозиции, образуя так называемую модальную рамку. В настоящее время существуют расхождения по вопросу о том, какие содержательные признаки следует относить к модусу. Ч. Филлмор считал принадлежащими модусу признаки отрицания, времени, наклонения и вида. Однако многие авторы значительно расширяют диапазон модуса, включая в его состав все компоненты предложения, не обусловленные семантической валентностью основного предиката, т. е. различные обстоятельственные компоненты. Соответственно этому в падежно-ролевой семантике нередко разграничиваются пропозициональные и модусные падежные роли.

Пропозициональные падежные роли. Пропозициональная роль представляет собой смысловое отношение аргумента к предикату. Поскольку предикаты различаются не только числом имеющихся у них аргументов, но и целым рядом других семантических признаков (Булыгина, Селиверстова, 1982), то характер падежных ролей, а в некоторой степени и их число будут зависеть от классификации предикатов. Такую зависимость можно обнаружить в концепциях У. Чейфа, У. Кука, Д. Нильсена, Р. Зимека, Т. Б. Алисовой и др. Например, у У. Чейфа (1975) предикаты действия непременно сопровождаются агентом, а предикаты состояния или процесса — пациентом. На наличие других падежных отношений влияют дополнительные избирательные единицы значения, характеризующие тот или иной класс предикатов.

В других случаях, например, в системах Ч. Филлмора, Ю. Д. Апресяна (1974), В. Г. Гака (1969), инвентарь падежных ролей прямо не связывается с типом предиката, однако неявно такая связь всегда присутствует.

Вопрос о числе падежных ролей является одним из самых сложных. Известны системы, включающие в себя только 4 роли (Шенк, 1980), и системы, в которых содержится 20 (Zimek, 1980) и даже 25 (Апресян, 1974) падежных ролей. Количество ролей зависит от многих факторов, среди которых можно назвать степень детализации смыслового отношения аргумента к предикату, учет или неучет лексики, заполняющей позиции аргументов, наличие специализированных формальных средств выражения ролей и т. д. Детализация означает, что одно и то же падежное отношение может быть представлено нерасчлененно или расчлененно. Например, у многих авторов (Ч. Филлмор, У. Чейф и др.) все значения, связанные с местоположением, включаются в единый локатив. Другие авторы (Ю. Д. Апресян) расчленяют такой глобальный локатив на значения места, исходной точки движения, конечной точки и маршрута. Если же использовать специализированные средства, имеющиеся в некоторых языках, например, в финно-угорских, то можно построить еще более дробную систему ло-

кативных значений. Так, в венгерском выделяется три падежа местонахождения: инессив, адессив и суперессив — и шесть падежей направления: иллатив, сублатив, аллатив, элатив, делатив и аблатив. Таким образом, все зависит от требований описания или от конкретно решаемой задачи. Для одной цели можно обойтись инвентарями с небольшим числом падежных ролей. В других случаях может потребоваться более тонкая система. Таким образом, верхний порог (максимальное количество падежных ролей) каждый автор определяет исходя из собственных целей. Что касается нижнего порога, то он обычно задается субъектно-объектным противопоставлением и меньше двух единиц содержать не может. Такие субъектно-объектные отношения называют иногда гиперролями и рассматривают в теории диатез как семантический уровень.

Смысловое содержание падежной роли зависит и от того, какие ограничения накладываются на лексику, заполняющую аргументную позицию. Например, одни лингвисты (ср. Ч. Филлмор) квалифицируют агенс как роль, которую может выполнять только одушевленное существительное, у других (например, у И. П. Сусова) функцию агенса может выполнять любое существительное. То же нередко относится и к другим ролям.

Модусные роли. Исследователи, выделяющие такие роли, указывают, что они относятся не к предикату, а ко всей пропозиции в целом. Фактически такие роли соответствуют содержательной характеристике того, что в синтаксисе принято называть обстоятельствами, сирконстантами или — в последнее время — детерминантами. Известно, что сирконстанты, в отличие от актантов, классифицируются по семантическому принципу (место, время, образ действия, причина и т. д.). Этот смысловой признак сирконстанта и есть то, что некоторые авторы квалифицируют как модусную роль. Например, У. Кук (Cook, 1979) к числу модусных ролей относит время, причину, цель, результат, образ действия, внешние (т. е. валентно не обусловленные) локатив и бенефактив.

В лингвистической литературе было показано, что сирконстанты, равно как и их смысловые признаки, называемые иногда модусными ролями, не являются компонентами пропозиции, а сами представляют собой свернутые пропозиции (Апресян, 1969). Было также отмечено, что явная семантическая выраженность сирконстантов, проявляющаяся в их семантической классификации, может рассматриваться как средство компенсации дефицита их валентной выраженности (Богданов, 1978).

Понятие модусной роли не получило широкого распространения, поэтому в центре внимания продолжают оставаться проблемы пропозициональных ролей.

Критерии реляционности и категориальности при формиро-

вании инвентаря падежных ролей. При определении состава инвентаря падежных ролей одни исследователи в большей мере склонны опираться на критерий реляционности, другие учитывают и категориальные признаки. Согласно критерию реляционности при выделении падежной роли необходимо учитывать прежде всего отношения аргументов друг к другу и к предикату. Например, различие между агенсом и пациенсом заключается, по мнению сторонников этой точки зрения, в том, что агенс воздействует, а пациенс подвергается воздействию. Обратное, при реляционном подходе, ни при каких случаях не может быть верным. Что же касается категориальных признаков слов, выполняющих функции агенса и пациенса, т. е., например, различие по признакам одушевленность — неодушевленность, лицо — нелицо и т. д., то оно считается в данном случае несущественным.

Критерия реляционности придерживаются И. П. Сусов, Ч. Филлмор (в последних вариантах своей концепции), Д. Нильсен, С. Староста. По мнению Ч. Филлмора, если бы мир состоял из объектов одного класса, например людей, то и на нем можно было бы задать все падежно-ролевые отношения, так как «один человек может схватить и поднять другого человека, воспользовавшись его телом как инструментом для того, чтобы сбить третьего человека, последний может быть ошеломлен и так далее. Короче говоря, в универсуме только с одним классом объектов легко представить себе падежные отношения агенса, инструмента, пациенса и испытывающего лица» (Филлмор, 1981, с. 508). Сказанное, конечно, верно в отношении мира, состоящего из таких максимально активных объектов, каковыми являются люди. Однако на мирах, состоящих из других объектов, например, только неодушевленных, задать многие отношения невозможно. Поэтому полностью отвлечься от категориальных признаков довольно трудно. На важность этих признаков указывали многие исследователи. Например, Р. Ван-Валин и У. Фоли в этой связи писали: «Семантическое отношение, которым ИГ (именная группа. — *В. Б.*) может быть связана со своим предикатом, ограничено его внутренним лексическим содержанием, например, агенсы и экспериенцеры являются обычно одушевленными, тогда как инструментальные и локативные ИГ — нет» (Ван-Валин, Фоли, 1982, с. 384).

Важным категориальным признаком, который обычно учитывается большинством исследователей, является признак предметности — непредметности аргумента. Функцию предметного аргумента выполняет существительное с вещным значением. Непредметным аргументом соответствуют пропозитивные существительные, в качестве которых выступают отпредикатные существительные, инфинитивы, герундии и т. д. Замена вещного существительного пропозитивным практически всегда

меняет характер падежно-ролевого отношения. Ср.: *Петр разбудил Ивана* (*Петр* — агенс) — *Взрыв разбудил Ивана* (*взрыв* — причина); *Иван открыл дверь ключом* (*ключ* — инструмент) — *Иван открыл дверь пинком* (*пинок* — действие) и т. д. Пропозитивное существительное в подобных примерах может быть развернуто в отдельную пропозицию. Ср.: *Что-то взорвалось, и это разбудило Ивана; Иван пнул дверь, и она открылась.* Такие развертки для случая предметных аргументов либо невозможны, либо чрезвычайно искусственны. Вообще наличие пропозитивных существительных в некотором предложении свидетельствует о его полипропозициональном строении (Богданов, 1977, с. 77—107).

Единичность семантической роли в пропозиции и множественность ролей в предложении. В составе пропозиции каждая падежная роль представлена только один раз. Принцип однократности роли был сформулирован Ч. Филлмором и остается верным для подавляющего большинства случаев. Исключение из этого правила составляют, пожалуй, только симметричные предикаты или предикаты, в состав которых входит семантический компонент симметричности. Ср.: *Иван похож на Петра* (два дескриптива, см. Богданов, 1977, с. 55); *Иван познакомился с Марией* (два агенса); *Петр познакомил Ивана с Марией* (*Петр* — агенс, *Иван* и *Мария* — пациенсы).

В составе предложения, которое семантически может представлять собой полипропозициональный комплекс, возможно не только несколько одинаковых ролей, но и наличие существительных, выполняющих одновременно несколько падежных ролей. Это связано с тем, что одно существительное может входить в состав нескольких пропозиций, а в предложении оно практически никогда не воспроизводится столько раз, в скольких пропозициях оно участвует. Например, в предложении *Старый матрос, уставший от тяжелой ноши, замедлил шаги и остановился* существительное *матрос* связано с четырьмя предикатами, так как участвует в четырех пропозициях, образующих данное предложение. По отношению к предикату *старый* оно выполняет падежную роль дескриптива, относительно предиката *уставший* — роль экспериенцера, а относительно предикатов *замедлил шаги* и *остановился* оно дважды выполняет функцию агенса. Таким образом, данное существительное совмещает в себе три разные падежные роли, причем одну из них дважды. В принципе количество выполняемых существительным ролей может быть значительно больше.

Имплицитное выражение аргументов и предикатов. Некоторые предикаты характеризуются тем, что включают аргумент с той или иной падежной ролью в состав своего лексического значения, благодаря чему этот аргумент не выступает в предложении в виде отдельной синтаксической позиции. Такое имплицитное выражение аргумента получило название

«кумуляции», или «минизации» (Почепцов, 1981). Ср.: *Кондуктор обилетил пассажиров* вместо *Кондуктор снабдил пассажиров билетами.* В первом предложении предикат *обилетить* кумулирует в своей семантике аргумент *билеты* с семантической ролью средства. В предложении *Корова отелилась* предикат кумулирует в себе аргумент *теленок* с ролью фактитива (результатива), ср.: *Корова родила теленка.*

Чаще всего кумуляции подвергаются аргументы с падежными ролями инструмента, средства, локатива и фактитива.

Аналогичное явление наблюдается и в сфере предикатов. Правда, в этом случае пропозиция выражается не в форме элементарного предложения, а чаще всего в виде именной группы, приобретающей в составе полипропозиционального предложения статус пресуппозиции. Ср.: *Деревянный дом рыбака стоял у самого моря.* В этом предложении имеются две пропозиции, а именно *Дом был построен из дерева* и *Дом принадлежал рыбаку*, выраженные именными группами *деревянный дом* и *дом рыбака,* причем предикат изготовления и предикат принадлежности представлены в этих группах имплицитно в составе обоих аргументов. Такое имплицитное выражение предикатов широко распространено практически во всех языках. Кумуляции способны подвергаться самые разные типы предикатов.

Кумуляцию, или минизацию, следует рассматривать как способ языкового кодирования семантики предложения. Иными словами, семантическое представление (пропозициональная структура), лежащее в основе двух предложений, а именно предложения с кумулятивным представлением аргумента или предиката и соответствующего ему предложения с обычным представлением, является тождественным, хотя указанные предложения и могут различаться стилистическими или иными коннотациями.

Явление, противоположное кумуляции, можно назвать развертыванием. Обычно оно выражается в том, что из предиката выделяется синтаксический актант, которому на пропозициональном уровне не соответствует никакой аргумент. Ср.: *Декан поблагодарил студентов — Декан выразил студентам благодарность; Германия воевала с Францией — Германия вела войну с Францией* и т. д. Таким двум вариантам предложения также соответствует одно пропозициональное представление, хотя они и могут иметь коннотативные различия.

Деривационные отношения в пропозиции. Пропозиции на близких по смыслу предикатах, различающихся числом аргументных позиций, могут быть связаны деривационно-семантическими отношениями. Под семантической деривацией можно понимать процесс выведения более сложных пропозициональных, а тем самым и предложенческих смыслов из более простых в форме некоторой последовательности шагов (Богданов,

1980). Обычно такая деривация связана с переводом предиката в другой семантический класс нередко с изменением числа его аргументов.

Проблемы семантической деривации довольно подробно анализировались У. Чейфом (1975, с. 139—166), а также и и другими авторами. По мнению У. Чейфа, деривация обусловлена существованием в сфере глаголов и существительных особых деривационных единиц, переводящих эти классы слов из одной категории в другую. Например, если к глаголу предложения *Дрова были сухи* добавить деривационную единицу «инхоатив», то глагол *быть сухим* из категории состояния переходит в категорию процесса, не изменяя при этом падежной роли своего единственного аргумента *дрова*, которые в обоих случаях соответствуют патиенту (пациенсу). Добавление же деривационной единицы «каузатив» к предикату (у У. Чейфа — глаголу) процесса *сохнуть* превращает его в предикат действия-процесса *сушить*, требующего второго аргумента с падежной ролью агента (агенса), т. е. в результате этой операции возникает предложение типа *Джон сушил дрова*, причем и здесь аргумент *дрова* продолжает оставаться в патиенте. Аналогичные явления У. Чейф усматривает и в сфере существительного, где слово из категории неисчисляемого может перейти в категорию исчисляемого, из единичного в множественное и т. д.

Учет деривационных отношений чрезвычайно важен, поскольку они позволяют связать предикаты, а следовательно, и целые пропозиции в единую и целостную систему.

Иерархизация падежных ролей. В падежно-ролевых концепциях не все роли, входящие в состав пропозиции, считаются тождественными. Обычно какие-то пропозициональные роли рассматриваются как более важные, тогда как другие отодвигаются на задний план. Все это приводит к различным видам иерархизации падежных ролей. В основу иерархизации может быть положен какой-то признак, например, признак деривационной важности падежа. Такую иерархизацию осуществляет, в частности, У. Чейф, который самыми важными считает патиент и агент. По его мнению, «категории патиент и агент носят фундаментальный характер, которым не обладают прочие отношения. В каждом предложении присутствует существительное-патиент или существительное-агент, или оба вместе, если только глагол не относится к особой амбиентной категории» (Чейф, 1975, с. 189).

Действительно, эти падежные роли имеют особый характер, так как каждая из них в отдельности участвует во фразонепроизводных структурах типа предикат — агенс (*Девочка прыгает*) или предикат — пациенс (*Дрова сухи*), а обе они образуют важнейшую структуру первой ступени производности предикат — агенс — пациенс (*Иван сушит дрова*); от этих

трех исходных структур образуются все производные предложения.

Иерархизация падежных ролей производится также с точки зрения выбора субъекта или перспективы. Это значит, что порядок ролей определяется преимущественным правом падежной роли на выполнение функции субъекта в предложении. Например, у Ч. Филлмора в статье «Дело о падеже» эта иерархия выглядела следующим образом: агентив — инструменталис — объектив. В работе «Некоторые проблемы падежной грамматики» указанная иерархия приобретает следующий вид: агент — экспериенцер — инструмент — объект — источник — цель. Следует иметь в виду, что во второй работе изменился и инвентарь падежей.

Субъектная иерархизация, называемая также субъективизацией или первичной топикализацией, интерпретируется иногда как выбор синтаксической перспективы, или такого изображения ситуации с помощью предложения, при котором ее наиболее важные компоненты попадают как бы в фокус зрения. Перспектива формируется правилами выбора прежде всего субъекта и во вторую очередь — прямого объекта. От нее в известной степени зависит и выбор предиката (Филлмор, 1981). В перспективу может попасть практически любой падеж.

Иерархизация падежных ролей производится также в системах У. Кука и С. Старосты.

Пропозициональная семантика как непадежного, так и падежно-ролевого типа продолжает интенсивно разрабатываться и в настоящее время, поскольку целый ряд ее положений и приемов имеет определяющее значение для всей синтаксической семантики в целом. Кроме того, существует масса проблем, весьма далеких от своего решения.

3.3. НЕПРОПОЗИЦИОНАЛЬНОЕ НАПРАВЛЕНИЕ

В рамках концептоцентрического подхода выделяются и такие теории, которые оперируют понятиями общей, или аристотелевской, логики. Структура предложения в этой логике подводится под модель суждения, т. е. описывается в терминах субъекта и предиката и соединяющей их связки. При этом предикат интерпретируется как признак, ассертивно приписываемый субъекту, а субъект — как носитель признака. Связка устанавливает отношение присущности признака его носителю.

Такой принцип реализован, например, в семантической концепции Н. Д. Арутюновой (1976б). Субъектно-предикатные отношения устанавливаются между элементами трех типов: онтологическими, которые именуются предметами; мыслительными, или понятиями; и языковыми, иначе — именами.

Мысль человека способна двигаться от любого из названных элементов к любому другому, в результате чего возникают девять отношений, называемых логико-синтаксическими началами, из которых рассматриваются четыре, а детально анализируются только два. Каждый логико-синтаксический тип выражается определенным набором синтаксических конструкций и сводим к схемам общей логики. Эти схемы имеют следующий вид:

1. Отношение характеризации: $С_{реф}$ — $П_{нереф}$.

Мысль движется здесь от предмета к понятию. Субъект считается референтным, или ориентированным на денотат, а предикат — нереферентным, т. е. ориентированным на сигнификат. Ср.: *Петр добр; Петр — врач.*

2. Отношение экзистенции: $С_{реф/нереф}$ — $С_{нереф}$.

Мысль движется от понятия к предмету. Субъектом является локализатор, который может быть референтен и нереферентен, а предикатом — указание на наличие внутри локализатора определенных объектов. Ср.: *В лесу водятся волки; В этом лесу водятся волки.*

3. Отношение идентификации: $С_{реф}$ — $П_{т}$ — $С_{реф}$.

Мысль движется от предмета к предмету. Предикат выражает отношение тождества и выступает как своего рода зеркало, в котором отражается субъект. Именно поэтому субъект в схеме представлен дважды. Субъект обладает свойством референтности. Ср.: *Старуху убил Родион Раскольников = Убийцей старухи был Родион Раскольников.*

Субъектно-предикатный принцип моделирования смысла предложения используется также в интерпретативной семантике (Катц, 1981). Содержательным *аналогом предложения* в этой теории является семантически проинтерпретированная глубинная структура, в качестве которой выступает либо базовый С-показатель, если описываемое им предложение является элементарным (ядерным), либо базис, если в основе предложения лежит комплекс базовых С-показателей (Хомский, 1972, с. 104—105, 120—122). Базовый С-показатель имеет, как известно, схему NP — PrP, т. е. состоит из именной составляющей, практически эквивалентной субъектной составляющей, или субъекту, и предикатной составляющей, т. е. предиката. При наличии базиса субъект выделяется в каждом из образующих его базовых С-показателей. Базовый С-показатель можно уподобить пропозиции, образованной одноместным предикатом, в качестве которого выступает предикатная составляющая, и единственным аргументом, каковым является именная составляющая.

Что касается семантической интерпретации, то эта операция относится не к способу представления смысла предложения, она используется для отграничения осмысленных предложений от бессмысленных.

Субъектно-предикатный способ моделирования предложения используется в концепции И. П. Сусова на так называемой предикационной ступени порождения смысла (Сусов, 1973, с. 51—93).

Данная модель находит широкое применение также в теории актуального членения, где субъектно-предикатное противопоставление нередко интерпретируется в терминах противопоставления темы и ремы. Следует отметить, однако, что тематизация, т. е. выделение темы и противопоставление ее реме, и субъективизация, т. е. выделение в предложении субъекта и противопоставление его предикату, представляют собой связанные, но не тождественные явления. Оба явления отражают коммуникативное распределение смысла в предложении. Однако тематизация ориентирована на текстовые условия предложения и в принципе, по крайней мере в некоторых языках, возможна без субъективизации. Ср.: ***Сыну*** *отец подарил велосипед;* ***Велосипед*** *отец подарил сыну.* Слова, выделенные в этих предложениях, соответствуют теме, но не являются субъектами. При этом тематизация может быть достигнута и субъективизацией. Ср.: ***Отец*** *подарил сыну велосипед;* ***Велосипед*** *был подарен сыну отцом;* ***Сын*** *получил в подарок от отца велосипед.* Выделенные слова являются здесь субъектами и темами.

Если противопоставление тема — рема связано с понятием степени новизны информации для слушающего и в большей степени зависит от текста, то противопоставление субъект — предикат представляет собой логическую операцию, связанную с выделением предмета суждения и приписываемого ему признака. Если имеется пропозиция на многоместном предикате типа *дарить* или *давать,* то в позицию субъекта может быть выдвинут любой из трех его аргументов. С. Д. Кацнельсон считает, что такое выдвижение связано с тематизацией аргумента. Однако далеко не все полагают, что субъект обязательно тематичен. В то же время, по-видимому, общепризнан тезис о том, что субъект более тесно связан с предикатом, нежели тема (или топик). Все сказанное дает основание не смешивать субъект с темой, хотя и позволяет находить точки пересечения между ними.

Рассмотренный выше концептоцентрический подход отличается наибольшим разнообразием теорий. Это обусловлено тем, что у таких теорий давняя и богатая логическая традиция. Кроме того. создатель любого варианта теории чувствует себя в меньшей мере связанным теми или иными онтологическими ограничениями. Хотя принцип онтологического воздержания, свойственный всему этому подходу, можно рассматривать как известный недостаток, поскольку отражательную функцию языка он оставляет на заднем плане, он обладает и некоторыми достоинствами. Главное из них в том, что исследователь осво-

бождает себя от необходимости изучать экстралингвистический мир со всеми его сложностями, не имеющими прямого отношения к языку, и сосредоточивает основное внимание именно на языке.

4. СИНТАКТИКОЦЕНТРИЧЕСКИЙ ПОДХОД

К анализу семантики предложения некоторые авторы подходят не со стороны отражаемой предложением онтологии или содержащейся в предложении концептуальной структуры, а со стороны синтаксической схемы, реализованной в предложении. В этом случае исследование начинается с выделения и определения структурных схем, их систематизации и инвентаризации. Все прочие аспекты содержания предложения подчиняются принципу структурной схемы, через призму которой рассматривается и сама сентенциональная семантика.

Синтактикоцентрический подход изложен в «Русской грамматике» 1980 г. в разделе, написанном Н. Ю. Шведовой (Русская грамматика, 1980, с. 83—98, 123—136).

Структурная схема определяется как «абстрактный синтаксический образец, по которому может быть построено отдельное минимальное относительно законченное предложение» (там же, с. 93). Каждой схеме присущи такие компоненты, как формальное устройство и семантика. К формальному устройству относится характер и количество символов классов слов, образующих схему. На этом основании выделено 9 однокомпонентных и 22 двухкомпонентные схемы. Схем с числом компонентов больше двух не постулируется. Семантика схемы определяется как то значение, которое присуще схеме в отвлечении от грамматических значений компонентов и правил их лексического наполнения. Например, схеме $N_1 — V_f$ приписывается следующая семантика: «отношение между субъектом и его признаком — действием или процессуальным состоянием», а схеме Praed — семантика «наличие состояния — бессубъектного или отнесенного к субъекту». Примерами первой схемы могут служить предложения *Лес шумит; Завод работает; Дети веселятся,* примерами второй — предложения типа *Темно; Сыро; Скучно; Радостно* и др.

Помимо понятия семантики структурной схемы вводится еще понятие семантической структуры предложения, которая интерпретируется как «то его языковое значение, которое создается взаимным действием семантики структурной схемы предложения и лексических значений заполнивших ее слов» (там же, с. 87). Компонентами семантической структуры предложения считаются предикативный признак, субъект и объект, причем объект относится не к структурной схеме, а к ее распространителям. Учитывается степень соответствия семантической структуры и формальной структуры, образованной структурной схе-

мой и распространителями. Все предложения, имеющие одну и ту же семантическую структуру, но разное грамматическое оформление, считаются связанными формально-семантическими отношениями.

Рассмотренный подход характеризуется тем, что он полностью оставляет в стороне онтологический уровень. Что касается концептуального уровня, то он намечен здесь лишь контурно. Вклад семантики структурной схемы в семантическую структуру предложения не вполне ясен, с одной стороны, ввиду расплывчатости смысла структурной схемы, а с другой — вследствие того, что предложения типа *Отец подарил сыну велосипед* подводятся под схему $N_1 — V_f$, т. е. описываются через бессмысленные предложения типа **Отец подарил.*

В силу принципа асимметрического дуализма одна и та же структурная схема с разным лексическим наполнением может оказаться неоднозначной, т. е. реализовать разные семантические структуры. Именно поэтому некоторые лингвисты с полным основанием ставят под сомнение существование чисто схемной семантики, полностью отвлеченной от лексического наполнения.

Все структурные схемы в рассматриваемой концепции получают одинаковый статус, т. е. среди них не выделяются изосемичные и неизосемичные схемы. Но даже если бы это было сделано, структурные схемы все-таки остались бы явлением идиоэтнического характера, в то время как семантика предложения в большей степени может претендовать на универсальность.

Все сказанное свидетельствует о том, что моделирование семантики предложения на базе его структурной схемы оказывается во многих отношениях малоудачной попыткой и не приводит к получению удовлетворительных результатов.

5. АНТРОПОЦЕНТРИЧЕСКИЙ ПОДХОД

5.1. СПЕЦИФИКА АНТРОПОЦЕНТРИЧЕСКОГО ПОДХОДА

Характерная черта всех рассмотренных выше теорий заключается в том, что содержание предложений исследуется в них в отвлечении от тех, кто этими предложениями пользуется, т. е. от коммуникантов. Такой принцип исследования основывается на известном тезисе Ф. де Соссюра о том, что язык надо изучать в самом себе и для себя, а следовательно, все, что входит в сферу внешней лингвистики, не должно интересовать лингвиста. В трансформационной грамматике область употребления, не относящаяся к компетенции, также считалась второстепенной.

В настоящее время ситуация коренным образом изменилась. Доказано, что без учета факторов, связанных с коммуникантами и с условиями коммуникации, т. е. условиями речевого акта,

семантика предложения не может быть удовлетворительно описана и не может получить однозначной интерпретации. Область исследований, связанную с коммуникантами и условиями коммуникации, принято называть лингвистической прагматикой. Лингвистическая прагматика получила в последнее время большое развитие.

5.2. ФАКТОР ГОВОРЯЩЕГО

Роль говорящего (или пишущего в случае письменной коммуникации) проявляется прежде всего в его коммуникативной интенции, которую можно определить как намерение произнести высказывание с определенной целью. Коммуникативное намерение говорящего реализуется в иллокутивной функции, являющейся своего рода инструкцией для слушающего, в которой указано, как интерпретировать пропозициональное содержание высказывания, произнесенного говорящим. Дело в том, что без такой инструкции пропозициональное содержание никогда не сообщается.

В настоящее время различные авторы приводят разные списки иллокутивных функций. Среди них чаще всего встречаются: 1) ассертивы — утверждения, констатации, сообщения; 2) директивы — приказы, требования, просьбы, советы; 3) декларативы — приговоры, объявления, благословения; 4) комиссивы — обязательства, заверения, гарантии, обещания; 5) экспрессивы — благодарности, поздравления, пожелания, извинения, сожаления и т. д. В пределах некоторых видов иллокутивных функций выделяются иногда подвиды. Например, директивы подразделяются на инъюнктивы — приказы, реквестивы — просьбы, инвитативы — приглашения. Некоторые авторы к директивам относят и вопросы.

Каждая иллокутивная функция выражает определенное коммуникативное намерение говорящего. Например, комиссив выражает предложение говорящего, чтобы слушающий положился на его обязательство создать то положение дел, которое выражено в пропозиции.

Для выражения иллокутивных функций существуют различные средства, называемые индикаторами иллокутивной силы, или аллокуциями. Среди них самым эксплицитным средством является перформативный глагол. Высказывание, содержащее перформативный глагол, принято называть перформативным высказыванием. Ему присущи следующие основные характеристики: 1) эквиакциональность, или равнозначность действию. Ср.: *Я обещаю тебе вернуть твою книгу* есть сам акт обещания, но *Я посещаю уроки музыки* есть сообщение о посещении; 2) неверифицируемость, т. е. неприложимость к таким высказываниям критериев истинности — ложности. Ср.: высказывание *Я обещаю тебе вернуть твою книгу* не является ни истинным, ни лож-

ным; 3) автореферентность, или способность такого высказывания отсылать к собственному речевому акту; 4) автономинативность, т. е. способность именовать собственный речевой акт; 5) эквитемпоральность, или совпадение момента действия, обозначаемого перформативным глаголом, с моментом речи; 6) компетентность, т. е. правомочность говорящего осуществить действие, обозначенное перформативным глаголом, ср.: *Объявляю собрание открытым;* 7) синсеративность, или условие искренности говорящего; 8) уникальность, т. е. неповторимость; 9) определенная грамматическая выраженность.

Не всякая иллокутивная функция имеет перформативное выражение. Например, для угрозы, оскорбления, критики и т. д. не существует перформативных высказываний.

Иллокутивная функция может выражаться и другими средствами. Среди них можно назвать интонацию, с помощью которой можно выразить просьбу, приказ, вопрос и т. д.; категорию наклонения (ср. повелительное наклонение для директива, изъявительное — для ассертива и т. д.); морфологические средства (ср. использование инфинитива в русском языке для выражения инъюнктива: *Встать*!); модальные слова, частицы, междометия и т. д. В отличие от перформативных высказываний указанные средства не обладают однозначностью, поэтому может потребоваться их дополнительное уточнение.

Иллокутивные функции характеризуют прежде всего говорящего, отражая его коммуникативное намерение, т. е. то, чего он хочет достичь с помощью речи. Однако из этого не следует, что говорящий не учитывает слушающего. Говорящий может успешно реализовать свое коммуникативное намерение только в том случае, если он должным образом учитывает особенности слушающего. А для этого он должен располагать некоторой моделью последнего.

5.3. МОДЕЛЬ СЛУШАЮЩЕГО

Модель слушающего включает в себя большое количество характеристик, которые в лингвистической литературе должным образом еще не систематизированы. Однако к числу наиболее важных характеристик, которые должны учитываться говорящим в первую очередь, если он заинтересован в успешности речевой коммуникации, относятся следующие: 1) языковая компетенция слушающего, т. е. знание им языка, на котором к нему обращается говорящий; 2) социальная и профессиональная характеристика; 3) национальная характеристика слушающего, в частности, культурные и этнические нормы того мира, к которому он принадлежит; 4) биологические данные: мужчина — женщина, взрослый — ребенок, здоров — болен, имеет — не имеет физических недостатков и т. д.; 5) психологический тип: циклотимический — шизотимический, интровертный — экстра-

вертный; темперамент: холерик, сангвиник, флегматик, меланхолик; 6) текущее психическое состояние; 7) вкусы и пристрастия слушающего; 8) семейное положение: женат (замужем) — холост (не замужем), имеет — не имеет детей, родителей, других родственников; 9) характер текущих целей и интересов; 10) степень знакомства говорящего со слушающим.

Ни один говорящий не располагает полной моделью слушающего. Однако при полном отсутствии у говорящего информации о слушающем речевая коммуникация может оказаться неэффективной.

Речевое воздействие говорящего на слушающего приводит к изменению внутреннего состояния последнего. Это изменение принято называть термином «перлокутивный эффект». Перлокутивный эффект может как сознательно планироваться говорящим, так и не планироваться. В любом случае он может достигаться различными компонентами речевого акта, а именно: 1) пропозициональным содержанием высказывания (ср. действие информации о неприятных для слушающего событиях); 2) типом иллокутивной функции (ср. реакцию на неправильно выбранную иллокутивную функцию: «Какое вы имеете право приказывать мне?»); 3) способами выражения того и другого (например, большое значение имеет выбор интонационно-мелодического рисунка, синтаксической конструкции, подбор лексики, *распределение старой и новой информации* и т. д.).

Ориентация говорящего на модель слушающего создает коммуникативное взаимодействие между ними, или то, что в последнее время принято называть кооперацией.

5.4. ПРИНЦИП КООПЕРАЦИИ В РЕЧЕВОМ ОБЩЕНИИ

Речевое общение регулируется определенным ритуалом, получившим название «принципа кооперации». Сущность его состоит в том, что коммуниканты, вступая в речевое общение, обычно преследуют цель достичь взаимопонимания и ведут себя соответственно этому, т. е. как бы кооперируются друг с другом.

Принцип кооперации реализуется в форме описанных Х. Грайсом (1985) коммуникативных постулатов, или конверсационных максим. К числу их относятся: 1) постулат количества, или информативности, определяющий количество информации, сообщаемой в тех или иных речевых условиях адресату; 2) постулат качества, или истинности, запрещающий сообщать ложную информацию; 3) постулат релевантности, определяющий уместность высказывания; 4) постулат способа, или ясности, требующий однозначности высказывания.

Указанные постулаты относятся к идеальному способу коммуникации, поэтому в реальном речевом общении никогда полностью не соблюдаются. Можно привести массу примеров, ког-

да они нарушаются. Например, постулат количества нарушается в случае тавтологических высказываний типа *Женщина есть женщина,* постулат качества — в случае иронии или гиперболы, постулат ясности — в случае каламбуров и т. д. То или иное отклонение говорящего от принципов кооперации, выраженных постулатами, может использоваться им преднамеренно и дает возможность слушающему извлечь из высказывания информацию, не имеющую отношения к его смыслу, а вытекающую только из постулатов. Такую не выраженную явно в высказывании информацию принято называть импликатурой. Одно и то же высказывание может иметь множество импликатур. Каждая импликатура определяется конкретными условиями общения. Например, импликатурами тавтологического предложения *Женщина есть женщина* могут быть следующие смыслы: 1) женщины слабы и беспомощны, 2) женщины сильны и выносливы, 3) женщины мелочны, тщеславны и завистливы, 4) женщины преданны, заботливы и внимательны — и т. д. практически до бесконечности.

Таким образом, включение предложения в прагматический контекст вносит необходимые уточнения в его содержание и превращает его в высказывание.

Антропоцентрический подход в определенной степени противоположен онтологоцентрическому, так как в первом случае содержание предложения ориентировано вовнутрь, т. е. на субъективный фактор, тогда как во втором случае оно ориентировано вовне, т. е. на объективный фактор, на внешний мир.

6. КОМПЛЕКСНЫЙ ПОДХОД

6.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Рассмотренные выше четыре основных подхода к моделированию содержания предложения характеризуются тем, что каждое из них акцентирует внимание на какой-то одной стороне и в соответствии с этим вырабатывает методику исследования, основные понятия и соответствующую терминологию. В связи с этим естественно возникает вопрос о том, достаточно ли для описания семантики предложения выбрать один из подходов и руководствоваться принятыми в нем процедурами анализа или целесообразно стремиться к комплексному анализу, учитывающему разнообразные факторы содержания. Современное состояние синтаксической семантики свидетельствует о том, что более перспективен комплексный подход. Новейшие исследования в этой области подтверждают такой вывод. Разумеется, комплексный подход не должен представлять собой эклектического объединения различных принципов и методик. Это значит, что внутри него должен выдерживаться некоторый приоритетный порядок. Он должен основываться на приоритете

принципов концептоцентрического анализа падежно-ролевого типа при учете онтологического фактора в той мере, в какой это необходимо. Это значит, что онтология должна учитываться как исходный и определяющий фактор для языка в целом, а не для каждого высказывания в отдельности. Язык в целом выполняет отражающую функцию, позволяя коммуникантам говорить о том, что находится вне языка. Язык, который был бы лишен этой способности, нельзя было бы назвать человеческим языком, так как он был бы лишен самого главного качества любого естественного языка. Но из этого вовсе не следует, что каждое высказывание должно описывать некоторый факт реальной онтологии. Дело в том, что онтологические корреляты для некоторых высказываний могут отсутствовать; для других высказываний и даже для целых классов их они могут оказаться неопределенными — ср., например, онтологический статус отрицательных высказываний (Бондаренко, 1983). Существует огромное количество текстов, например беллетристических, для которых вопрос об онтологии их высказываний также представляется непростым. В то же время независимо от характера онтологической соотнесенности содержание предложения может быть выражено концептуальной структурой, допускающей тот или иной вид репрезентации.

На каждую концептуальную структуру, лежащую в основе предложения, неизбежно накладываются прагматические характеристики, *корректирующие или видоизменяющие пропозициональное* содержание высказывания. В этом и состоит явление актуализации. Следовательно, прагматические параметры также должны найти отражение в модели содержания высказывания. Что касается формально-синтаксического аспекта, то, естественно, и он не может быть оставлен вне поля зрения, однако он должен занимать подчиненное место в иерархии содержательных приоритетов, поскольку он является всего лишь способом *выведения содержания высказывания на поверхность*, т. е. способом его реализации, или манифестации.

Четыре рассмотренных выше подхода, равно как и комплексный подход, о котором ниже пойдет речь, — это всего лишь приемы моделирования языковых, а не энциклопедических знаний, для которых существуют другие способы представления.

6.2. СЕМАНТИКА ЭЛЕМЕНТАРНОГО ПРЕДЛОЖЕНИЯ

В лингвистической литературе широко распространена точка зрения о номинативном характере предложения. Обычно под этим имеют в виду тот факт, что в предложении запечатлен образ называемой им ситуации. Этот смысловой образ ситуации принято называть термином «пропозиция».

Между ситуацией и пропозицией существует отношение пол-

ного структурного изоморфизма. Это значит, что признаку ситуации соответствует предикат пропозиции, а партициианту ситуации — аргумент пропозиции. Количество и характер партиципантов ситуации полностью соответствуют количеству и характеру аргументов пропозиции. Под характером здесь понимаются роли партиципантов в ситуации, которые полностью совпадают с ролевыми, или глубинно-падежными характеристиками аргументов в пропозиции. Различие между ситуацией и пропозицией состоит в том, что первая есть онтологический объект, а вторая — мыслительный, или логический, феномен.

Предикаты пропозиции могут быть расклассифицированы по числу мест. Каждый из таких классов будет соответствовать определенному типу онтологического признака. Так, нуль-местные предикаты соответствуют признакам, присущим метеорологическим или временным ситуациям. Одноместный предикат выражает однопартиципантные свойства, состояния, процессы или действия. Двухместный предикат может быть либо выразителем чистого отношения, например отношения принадлежности, либо отношения с признаком акциональности или процессуальности. В случае трехместных предикатов в пропозиции выражается несколько отношений.

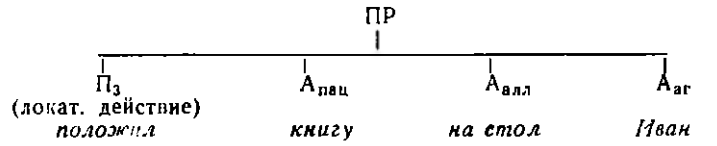
Число аргументов в пропозиции зависит от местности предиката, а семантический характер аргументов — от типа предиката, т. е. от того, выражает ли он свойство, состояние, действие и т. д. Ср., например, для одноместных предикатов: *Мальчик прыгает,* где *прыгает* — действие и, следовательно, аргумент *мальчик* является агенсом, но в случае *Мальчик болеет* предикат *болеет* выражает состояние и, следовательно, *мальчик* является не агенсом, а пациенсом.

То, что пропозиция изоморфна ситуации, не должно вызывать удивления, так как пропозиция есть результат логического отражения ситуации в предложении. При отражении природа отражаемого и отражающего будет различной, а структура тождественной. Если бы указанная закономерность не имела места, пришлось бы предположить, что логика отражает онтологию наподобие кривого зеркала. Из сказанного, однако, вовсе не следует, будто отношение изорморфизма сохраняется также и между пропозицией и предложением (элементарным). Такое отношение в принципе может быть, но может и не быть. Оно наблюдается в случае ядерного воплощения элементарного предложения. В остальных случаях наблюдается большее или меньшее несоответствие структур пропозиции и предложения. Элементарное предложение отличается от всех других видов предложений тем, что в нем реализована одна пропозиция (или одна атомарная пропозиция, если пользоваться терминологией логики).

6.3. НЕПРЕДЛОЖЕНЧЕСКИЕ СПОСОБЫ ВЫРАЖЕНИЯ ПРОПОЗИЦИИ

Пропозиция как отражение ситуации в принципе может быть выражена не только с помощью элементарного предложения, но и другими, непредложенческими, способами или синтаксическими структурами непредложенческого типа. Среди них можно назвать именные группы, инфинитивные, причастные, деепричастные, герундиальные и прочие конструкции. Если взять в качестве примера элементарное предложение в ядерном варианте *Ребенок плачет* и именную группу *Плач ребенка,* то обе эти синтаксические конструкции не различаются с точки зрения выражаемой ими пропозиции, которая здесь является одной и той же. Однако между этими конструкциями имеется и различие. Оно заключается в том, что приведенное предложение выражает пропозицию изоморфно, так как предикату соответствует финитный глагол, а единственному аргументу — существительное в форме подлежащего. В именной группе такого изоморфизма не наблюдается, так как и предикат и аргумент выражаются существительными, хотя и разными, поскольку одно из них является вещным существительным, а другое — отглагольным, т. е. транспонированным. Возможно и такое выражение пропозиции, когда какой-то компонент ее, т. е. предикат или аргумент, на синтаксическом уровне никак не выражен, а либо вытекает из морфологии соответствующей конструкции, либо восполняется из окружающего текста. Ср.: *дом отца* (не выражен предикат принадлежности), *чтение книги* (не выражен аргумент, осуществляющий чтение) и т. д.

Пропозиция, независимо от синтаксического способа выражения, будет в дальнейшем репрезентироваться схемами следующего вида:

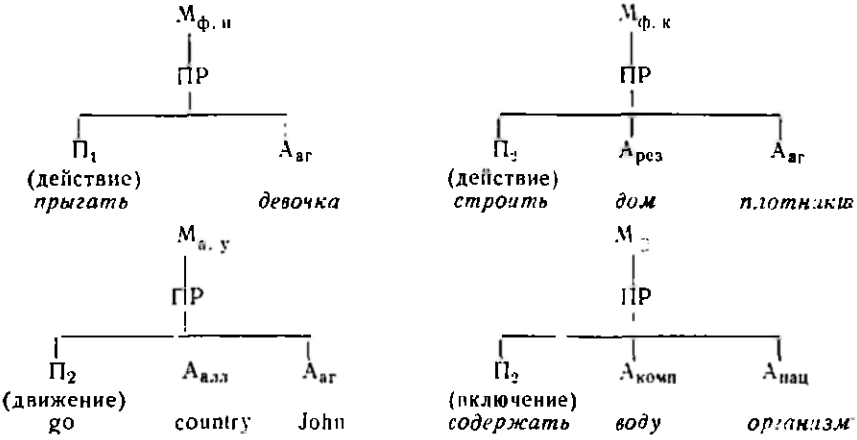


Данная схема выражает пропозицию, заключенную в предложении *Иван положил книгу на стол,* а также в непредложенческих комбинациях типа *Иван, положивший книгу на стол (сел и задумался); Иван, положив книгу на стол (вышел из комнаты); Ивану положить книгу на стол (было некогда)* и т. д. В приведенной схеме ПР означает пропозицию, $П_з$ — трехместный предикат локативного действия, $А_{пац}$ — аргумент со значением неодушевленного пациенса (пациентив, или пациент), $А_{алл}$ — аргумент со значением конечной точки (аллатив), $А_{аг}$ — аргумент со значением агенса (агентив).

6.4. МОДИФИКАЦИЯ ПРОПОЗИЦИИ

Пропозиция в составе реального предложения редко встречается в немодифицированном виде. Обычно она подвергается той или иной модификации. Средством модификации служит оператор с модифицирующей функцией, который можно называть модификатором. Поскольку сферой действия модификатора является единичная пропозиция, то модификатор фактически является одноместным предикатом высшего порядка.

Модифицирующую функцию по отношению к пропозиции выполняют операторы: 1) отрицания, 2) фазисности, 3) аспектуальности, 4) модальности, 5) каузативности. Модификация содержательно сводится к тому, что оператор либо меняет истинностное значение пропозиции на противоположное (отрицание), либо ограничивает действие предиката начальной, конечной или срединной фазой (фазисность), либо придает ему значение однократности, повторяемости, распределенности и т. д. (аспектуальность) и т. д. Единственным аргументом модификатора является вся пропозиция, поэтому в схемах они будут располагаться над пропозицией. Ниже приводятся примеры модифицированных пропозиций:

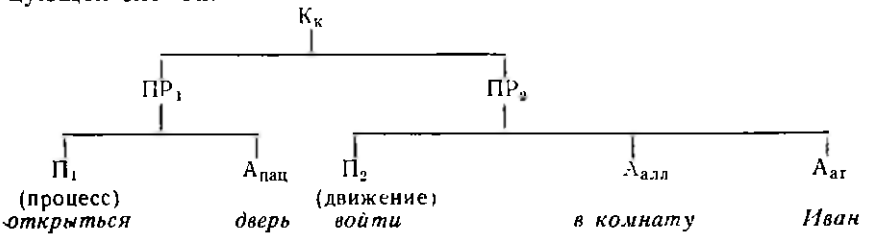


Условные обозначения: $М_{ф.\,н}$ — модификатор фазисный (начало), $М_{ф.\,к}$ — модификатор фазисный (конец), $М_{а.\,у}$ — модификатор аспектуальный (узитативность), М —модификатор модальный) (необходимость), $А_{рез}$, $А_{комп}$ — аргументы со значениями соответственно результата (результатив) и состава (композитив). Приведенные схемы описывают соответственно содержание предложений: 1) *Девочка начала прыгать* (или *запрыгала*), 2) *Плотники кончили строить дом*, 3) John used to go to the country, 4) *Организм должен содержать воду*. Пропозиция может быть подвергнута одновременно действию нескольких модификаторов. Ср.: *Иван не хотел начинать работу*, где представлены модификаторы отрицания, оптативной мо-

дальности и фазисности. Таким образом, модифицироваться может не только простая пропозиция, но и уже модифицированная пропозиция.

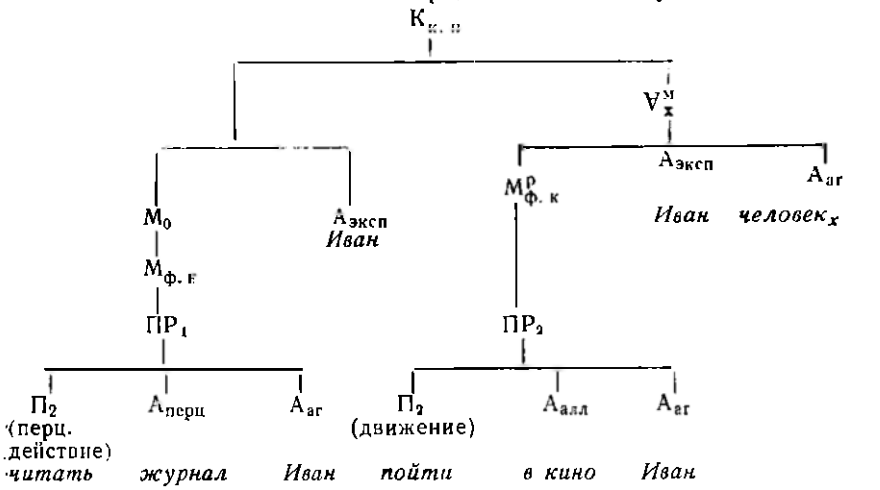
6.5. СОЕДИНЕНИЕ ПРОПОЗИЦИЙ И ОБРАЗОВАНИЕ ПОЛИПРОПОЗИЦИОНАЛЬНЫХ КОМПЛЕКСОВ

Отдельные пропозиции могут соединяться друг с другом с помощью соединительных операторов, которые в дальнейшем будут называться коннекторами. В отличие от модификатора коннектор представляет собой двухместный оператор, сферой действия которого являются минимум две пропозиции. К числу коннекторов относятся изучаемые в математической логике операторы конъюнкции, дизъюнкции, импликации, эквивалентности. Сюда относятся также коннекторы таксиса, причины/следствия, уступки и др. Например, предложение *Дверь открылась, и в комнату вошел Иван* может быть представлено следующей схемой:



Условные обозначения: $К_к$ — оператор конъюнкции, $ПР_1$ и $ПР_2$ — первая и вторая пропозиции.

Соединяться с помощью коннекторов могут также модифицированные пропозиции. Например, содержание предложения *Иван хотел приступить к чтению журнала, но все уговорили его пойти в кино* может быть представлено следующей схемой:



Условные обозначения: $К_{к.п}$ — коннектор конъюнкции со значением противительности, $М_о$ — модификатор оптативности, $V_x^м$ — квантор общности на ограниченном множестве (имеются в виду все люди, присутствовавшие в определенном месте), $М^р_{ф.к}$ — модификатор речевой фактитивной каузации, $А_{эксп}$, $А_{перц}$ — аргументы со значением соответственно лица, испытывающего некоторое состояние (экспериенсив), и объекта восприятия (перцептив), $человек_x$ — переменная x с областью допустимых значений, ограниченных значением «человек».

Использование коннекторов и модификаторов дает возможность рассматривать семантику достаточно сложных предложений как полипропозициональные комплексы. Важно при этом отметить, что некоторые лексемы, заполняющие в этих комплексах позиции некоторых аргументов, могут в принципе быть носителями нескольких ролевых отношений. Это зависит от того, в комбинацию с какими предикатами вступают соответствующие аргументы, ибо каждый аргумент получает свою ролевую характеристику по предикату. Так, в приведенном предложении лексема *Иван* дважды выступает в падежно-ролевом отношении агенса соответственно при предикатах *читать* и *пойти* и дважды в ролевом отношении экспериенсива соответственно при оптативном модификаторе *хотел* и при каузативном модификаторе *уговорили*. При использовании других непредложенческих способов выражения пропозиций количество лексем, выполняющих одновременно несколько падежных ролей, может быть значительно больше. Наличие модифицированных полипропозициональных комплексов с множеством ролей, совмещенных в одной лексеме, является свидетельством особой сложности смыслового строения предложения. Однако дело не исчерпывается только этим.

6.6. ДЕНОНАТИВНЫЙ (РЕФЕРЕНЦИАЛЬНЫЙ) СТАТУС ПРОПОЗИЦИЙ

Когда предложение содержательно соответствует комплексу пропозиций, то не каждая пропозиция, входящая в состав этого комплекса, выражается формой элементарного предложения. Чаще всего одна, реже — две или более пропозиций реализуются в форме предложения, тогда как остальные манифестируются именными группами, инфинитивными, причастными или деепричастными конструкциями или же придаточными предложениями. При осложнении пропозиций операторами предикаты также могут иметь форму нефинитных глаголов, а операторы могут выражаться как финитные глаголы. Так что фактическая картина может оказаться очень сложной. Однако пропозицию, выраженную формой элементарного предложения, есть основания считать главной, или матричной, а все остальные пропозиции — зависимыми, или подчиненными.

Со способом синтаксического выражения пропозиции теснейшим образом связан ее логический статус. Под логическим, денотативным или референциальным статусом пропозиции будем понимать ее отношение к критерию истинности/ложности. Поскольку верификация может быть связана с обращением к экстралингвистической действительности, то логический статус называют иногда денотативным. Принято выделять по крайней мере три денотативных статуса пропозиций: 1) ассертивный статус (статус утверждаемой пропозиции), 2) пресуппозиционный статус (статус заранее предполагаемой истины), 3) нейтральный статус (статус снятой утвердительности, т. е. статус пропозиции с неопределенным истинностным значением).

Ассертивный статус означает, что пропозиция подвергается операции утверждения или отрицания. Пропозицию, которая утверждается или отрицается, принято называть суждением. Как правило, суждения имеют форму самостоятельных повествовательных предложений. Часто достаточно одной лишь формы повествовательного предложения, чтобы квалифицировать его как суждение.

Однако пропозиция может выражаться не только формой самостоятельного предложения. Она может иметь форму именной группы, инфинитивной или какой-либо иной конструкции. Например, в предложении *Бег спортсмена радовал тренера* пропозиция с предикатом *радовал* утверждается, а пропозиция *Спортсмен бежал,* выраженная именной группой *бег спортсмена,* не утверждается. Она преподносится в форме самоочевидной истины, не требующей утверждения. Если бы этого не было, все предложение было бы лишено смысла.

Характерной особенностью таких пропозиций является то, что их истинность не меняется от того, утверждаем мы или отрицаем все предложение в целом. Ср.: *Бег спортсмена радовал тренера,* где пропозиция *Спортсмен бежал* остается истинной. Ср. также: *Бег спортсмена не радовал тренера,* где пропозиция *Спортсмен бежал* остается неизменной. В случае ассертивного статуса такое положение дел невозможно. Статус пропозиции, остающейся истинной независимо от истинности или ложности всего предложения, принято называть пресуппозиционным. Пропозиции с пресуппозиционным статусом вводятся с помощью так называемых фактивных предикатов. К их числу относятся предикаты *радовать(ся), знать, понимать, сожалеть, огорчаться, забывать, возмущать(ся), беспокоить(ся), удивлять(ся), быть странным, быть интересным, быть удивительным* и др. Пресуппозиции в данном случае выражаются формой именных групп или придаточных предложений.

В тех случаях, когда истинностное значение пропозиции остается неопределенным, имеет место нейтральный статус пропозиции. Пресуппозиция с этим статусом вводится с помощью нефактивных предикатов: *допускать, полагать, предполагать,*

думать, утверждать, считать, заявлять, быть возможным, быть вероятным и т. д.

Близка к пресуппозиции импликация. В отличие от пресуппозиции истинностное значение пропозиции при импликации подвержено действию отрицания, т. е. меняется при отрицании. Ср.: *Иван начал работать* — импликация: *Иван работает; Иван не начал работать* — импликация: *Иван не работает; Ивану удалось поступить в институт* — импликация: *Иван поступил в институт; Ивану не удалось поступить в институт* — импликация: *Иван не поступил в институт.*

Пропозиция всегда реализуется в каком-либо одном статусе. Никогда не бывает так, чтобы пропозиция одновременно была суждением и пресуппозицией.

На истинностное значение пропозиций оказывают влияние операторы — модификаторы и коннекторы. Например, в предложении *Иван хотел приступить к чтению журнала, но все уговорили его пойти в кино* модификатор начальной фазы *приступить* действует в форме прошедшего времени как импликация, т. е. предложение *Иван приступил к чтению журнала* означает, что *Иван читал журнал.* Однако оптативный модификатор *хотел* меняет статус данной пропозиции на нейтральный. Наличие же противительного союза *но* снимает эту нейтральность и превращает ее в пресуппозицию, в результате чего предложение *Иван хотел приступить к чтению журнала* воспринимается в смысле *Иван не читал журнала.* С другой стороны, каузативный модификатор *уговорили* во второй части предложения *Все уговорили его пойти в кино* действует как импликация, в результате чего пропозиция *Иван пошел в кино* воспринимается как истинная. На истинностное значение пропозиций действуют также грамматические значения, роль которых будет рассмотрена ниже.

Денотативный статус пропозиции должен найти свое отражение в семантической структуре предложения. Соответствующие обозначения будут использоваться и в схемах, репрезентирующих семантику предложения.

6.7. ВКЛЮЧЕНИЕ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ В ПРАГМАТИЧЕСКУЮ РАМКУ

Семантика элементарного предложения рассматривалась выше как одна пропозиция, а смысловое содержание более сложных предложений интерпретировалось как комплекс пропозиций. Однако фактически до сих пор мы имели дело не с пропозициями, а с так называемыми пропозициональными формами. Пропозициональная форма отличается от пропозиции тем, что аргументами первой являются переменные, а аргументами второй — константы. Типичным способом выражения переменных в языке являются общие имена, или имена нарица-

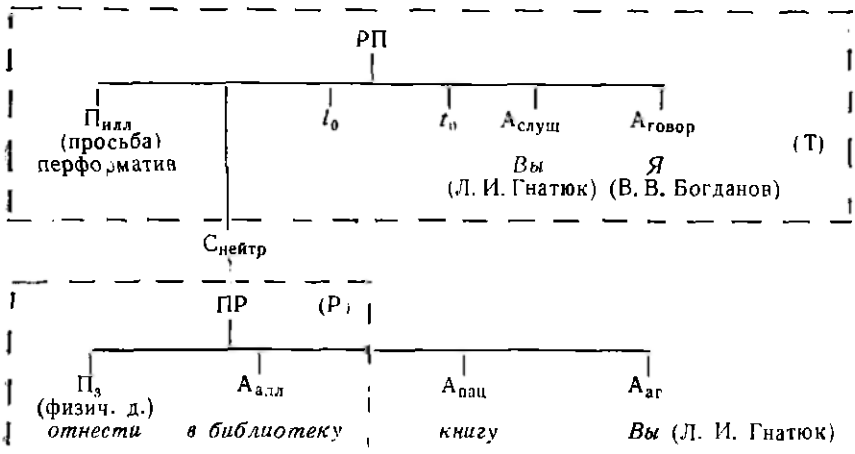
тельные. Например, в предложении *Мальчик идет в школу* такими именами являются *мальчик* и *школа*. Каждая из этих имен-переменных имеет область допустимых значений, определяемых их лексическими значениями. Поэтому такое имя является не обычной свободной переменной, которая может принимать любые значения, а переменной с ограниченной областью допустимых значений. То, что это переменная, доказывается тем, что словом *мальчик* обозначен здесь не какой-то конкретный и единичный мальчик, а практически любое человеческое существо, по отношению к которому может быть применено имя *мальчик*. То же самое в принципе справедливо и по отношению к слову *школа*. На уровне предложения не имеется в виду какая-то конкретная школа.

Поскольку *мальчик* и *школа* являются в данном предложении переменными, то, естественно, не может возникать и вопрос об истинности или ложности данного предложения. Данное предложение не имеет истинностного значения. Собственно говоря, смыслом предложения является либо одна пропозициональная форма, либо комплекс таких форм. Чтобы пропозициональную форму превратить в нормальную пропозицию, которая может быть истинной или ложной, предложение необходимо поместить в прагматический контекст, или в прагматическую рамку, т. е. использовать его в конкретном речевом акте. Это предполагает, что здесь вполне конкретными будут говорящий, слушающий, время и место речевого акта. Иначе говоря, предложение попадает в зону прагматических констант. В таких условиях предложение референциально привязывается к конкретной ситуации, или актуализируется, и тем самым превращается в высказывание. А это приводит к тому, что переменные пропозициональной формы заменяются константами пропозиции. В результате слово *мальчик* в приведенном выше предложении в составе высказывания будет обозначать конкретного, единичного мальчика, а слово *школа* — конкретную школу. В языке имеются специальные средства, которые позволяют говорящему производить актуализацию. Их именуют актуализаторами, и к их числу относят преимущественно местоимения различных классов (Падучева, 1985).

Разумеется, существуют предложения, содержание которых не зависит от условий речевого акта. Они референтны на уровне языка, следовательно, их аргументами являются константы, а их содержание соответствует пропозициям, а не пропозициональным формам. Здесь имеются в виду предложения типа *Волга впадает в Каспийское море; Земля вращается вокруг Солнца; Наполеон Бонапарт умер в 1821 году* и т. д. Эти предложения можно назвать предложениями-высказываниями.

Прагматическая рамка помимо тех прагматических констант, которые были названы, должна включать в себя информацию о конкретной иллокутивной функции, которая выражa-

ет коммуникативное намерение говорящего. Если эта функция репрезентирована перформативным глаголом, то данный глагол также должен найти отражение в семантическом представлении высказывания. В качестве примера рассмотрим смысловое представление высказывания *Я прошу вас отнести книгу в библиотеку.*



Условные обозначения: РП — рамка прагматическая, $П_{илл}$ — иллокутивный предикат просьбы в перформативном представлении, $А_{говор}$, $А_{слуш}$ — аргументы прагматической рамки, а именно говорящий и слушающий, с указанием, кто конкретно имеется в виду; t_0 — время речевого акта просьбы, конкретно 28 июня 1984 г. в 14.00; l_0 — место речевого акта — конкретно кафедра математической лингвистики ЛГУ; $С_{нейтр}$ — нейтральный статус пропозиции, поскольку перформативный глагол всегда вводит пропозицию с нейтральным статусом; (Т) — тематический фрагмент высказывания, (Р) — рематический фрагмент высказывания; библиотека = библиотека им. А. М. Горького ЛГУ, книга = Syntax and Semantics. Vol. 3. Speech Acts.

Однако иллокутивная функция не во всех случаях выражается перформативным глаголом. В такой ситуации прагматическая рамка формируется иными средствами, среди которых немалую роль играют грамматические категории, ориентированные на прагматику.

6.8. РОЛЬ ГРАММАТИЧЕСКИХ КАТЕГОРИЙ В ФОРМИРОВАНИИ СМЫСЛА ВЫСКАЗЫВАНИЯ

Грамматические категории, реализующиеся в высказывании в виде тех или иных грамматических значений, небезразличны для его содержания. Во многих случаях эти значения шифруют какие-то важные смысловые характеристики высказывания.

Поскольку в составе высказывания есть то, что относится к семантике, а именно пропозициональное содержание высказывания, и есть то, что относится к прагматике, а именно прагматическая рамка, то и в грамматических категориях наблюдается это деление, а именно, одни грамматические категории ориентированы на семантику, а другие — на прагматику. Грамматические категории, ориентированные на семантику высказывания, будем называть не-шифтерными категориями (Якобсон, 1972). К их числу относятся род, число, залог, вид, таксис, причем род, число и вид являются десигнаторами, а залог и таксис — коннекторами. Различие между ними заключается в том, что десигнаторы характеризуют либо только предикат пропозиции (как, например, вид), либо только аргументы пропозиции (как, например, род и число). Коннекторы же устанавливают связь между аргументами и предикатом одной пропозиции (залог) или между предикатами разных пропозиций. Десигнаторы сплошь и рядом выполняют также функцию модификаторов пропозиции.

Грамматические категории, ориентированные на прагматику высказывания, т. е. на установление связи между его пропозициональным содержанием и содержанием его прагматической рамки, будем называть шифтерными категориями. К их числу относятся лицо, время, наклонение и засвидетельствованность, причем первые две являются десигнаторами, а две вторые — коннекторами. Например, лицо соотносит участников прагматической рамки, т. е. коммуникантов, с аргументами пропозиции. Ср.: *Я читаю книгу=Я сообщаю вам, что я читаю книгу,* где аргумент прагматической рамки, т. е. говорящий, совпадает с аргументом пропозиции, т. е. субъектом чтения. Разумеется, прагматическая рамка здесь не представлена перформативным глаголом со своими аргументами, а присутствует в содержании приведенного высказывания имплицитно, что создает особую ситуацию, так как такое высказывание уже не является перформативным, а значит, его денотативный статус будет иным.

Такой десигнатор, как время, соотносит время прагматической рамки (момент речи) с временем пропозиции. Ср.: *Студенты пошли в кино = Я сообщаю вам, что студенты пошли в кино,* где время ухода студентов предшествует времени сообщения об этом.

Наклонение связывает говорящего прагматической рамки с пропозициональной частью, т. е. наклонение показывает, как говорящий оценивает связь между предикатом и аргументами пропозиции. Например, в предложении *Студенты пошли в кино* (изъяв. накл.) говорящий считает, что связь между предикатом *пошли* и аргументами *студенты* и *кино* имела место в действительности. В случае высказывания *Студенты пошли бы в ки-*

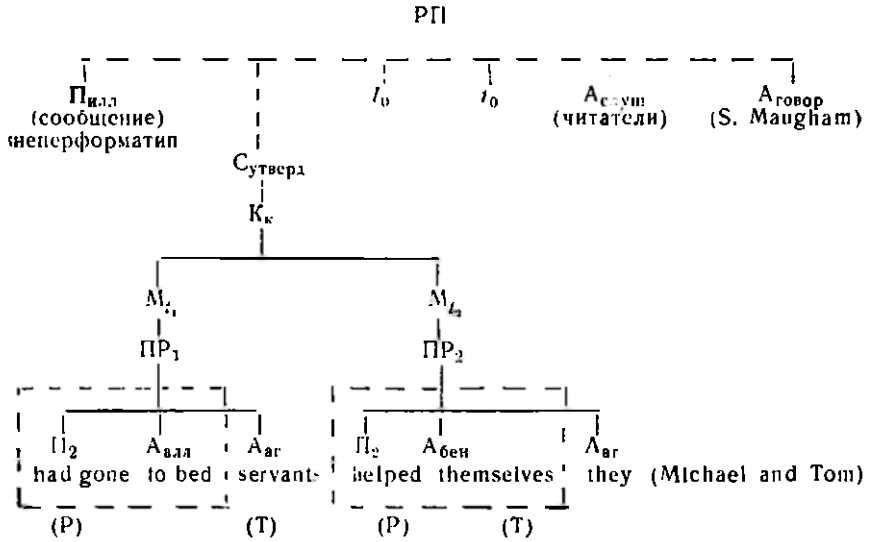
но (сосл. накл.) указанная связь оценивается говорящим как не имевшая места.

Засвидетельствованность связывает говорящего с пропозициональной частью с точки зрения источника пропозициональной информации. Иными словами, этот коннектор показывает, принадлежит ли пропозициональная информация самому говорящему или передается им со слов другого лица. Ср. болг.: *Студентите отивали на екскурзия=Говорят, что студенты отправились на экскурсию.*

Разумеется, в языке возможно и такое использование грамматических категорий, когда они в основном могут обслуживать потребности синтаксиса. К числу такого их употребления можно отнести использование прилагательными русского языка категорий рода, числа и падежа или использование глаголами категорий числа или рода (последнее в прошедшем времени). В данном случае эти категории выполняют в основном формально-согласовательную функцию. Правда, и здесь полностью семантическую функцию исключить нельзя. Действительно, в случае согласования глагола с подлежащим, а в других языках, как в венгерском, — и с прямым дополнением, а в дагестанских — и со всеми актантами, сам факт согласования по тем или иным грамматическим категориям выступает как показатель того, что данные существительные как аргументы относятся к данному глаголу как предикату. В случае согласования существительных с прилагательными в русском языке это тоже служит для указания на то, что данные существительные как аргументы относятся к данным прилагательным как предикатам при реализации пропозиции в пресуппозиционном статусе. Все это лишний раз свидетельствует о том, что грамматика в языке никогда не используется вхолостую, а всегда нагружена семантически или прагматически.

В заключение приведем схему (см. с. 199) семантической структуры высказывания без явно выраженной прагматической рамки. В качестве примера возьмем английское предложение The servants had gone to bed and they helped themselves (S. Maugham. Theatre).

Условные обозначения: пунктиром обозначена имплицитная, т. е. явно не выраженная в данном предложении прагматическая рамка, информация для которой берется из контекста всего произведения; говорящим в данном случае является автор произведения, т. е. С. Моэм, а слушающими — читатели; пропозициональная часть предложения представляет собой конъюнкцию двух пропозиций, причем обе пропозиции имеют утвердительный денотативный статус (ассертивный статус); M_{t_0} — время пропозиции $ПР_1$, M_{t_1} — время второй пропозиции $ПР_2$, причем $M_{t_1}<M_{t_2}$, и это отношение выражено категорией таксиса (сопоставлением времен Past Perfect и Past Indefinite). При этом $t_1<t_0$ и $t_2<t_0$, и эти отношения выражаются катего-



рией грамматического времени. Следует также отметить, что бенефициативный и агентивный аргументы пропозиции $ПР_2$ соответствуют одному референту, т. е. кореферентны.

Литература

Адамец П. Образование предложений из пропозиций. Прага, 1978.

Апресян Ю. Д. 1) Синтаксис и семантика в синтаксическом описании // Единицы разных уровней грамматического строя языка и их взаимодействие. М., 1969; 2) Лексическая семантика. М., 1974.

Арутюнова Н. Д. 1) Понятие пропозиции в логике и лингвистике // Изв. АН СССР. Сер. лит. и яз. 1976а. № 1, т. 35; 2) Предложение и его смысл. М., 1976б.

Богданов В. В. 1) Актанты и сирконстанты // Проблемы членов предложения в индоевропейских языках. Грозный, 1978; 2) Семантико-синтаксическая организация предложения. Л., 1977; 3) Механизм семантической деривации // Семантико-синтаксическая организация предложения и текста. Грозный, 1980; 4) Речевое общение: Прагматические и семантические аспекты. Л., 1990; 5) Текст и текстовое общение. СПб., 1993.

Бондаренко В. Н. Отрицание как логико-грамматическая категория. М., 1983.

Булыгина Т. В., Селиверстова О. Н., Елисеева А. Г. и др. Семантические типы предикатов. М., 1982.

Ван Валин Р., Фоли У. Референциально-ролевая грамматика // Новое в зарубежной лингвистике. Вып. XI: Современные синтаксические теории в американской лингвистике. М., 1982.

Винокур Т. Г. Говорящий и слушающий. М., 1993.

Витгенштейн Л. Логико-философский трактат. М., 1958.

Гак В. Г. 1) К проблеме синтаксической семантики (семантическая интерпретация «глубинных» и «поверхностных» структур) // Инвариантные синтаксические значения и структура предложения. М., 1969; 2) Высказывание и ситуация // Проблемы структурной лингвистики. 1972. М., 1973.

Грайс Г. П. Логика и речевое общение // Новое в зарубежной лингвистике. Вып. 16: Лингвистическая прагматика. М., 1985.

Катенина Т. Е., Рудой В. И. Лингвистическое знание в Древней Индии // История лингвистических учений: Древний мир. Л., 1980.

Катц Дж. Семантическая теория // Новое в зарубежной лингвистике. Вып. X: Лингвистическая семантика. М., 1981.

Лакофф Дж. О порождающей семантике // Новое в зарубежной лингвистике. Вып. X: Лингвистическая семантика. М., 1981.

Лингвистика на исходе XX века: итоги и перспективы: В 2 т. М., 1995.

Макколи Дж. О месте семантики в грамматике языка // Новое в зарубежной лингвистике. Вып. X: Лингвистическая семантика. М., 1981.

Падучева Е. В. 1) Высказывание и его соотнесенность с действительностью. М., 1985; 2) О семантике синтаксиса. М., 1974.

Панкрац Ю. Г. Пропозициональные структуры и их роль в формировании языковых единиц разных уровней. Минск; Москва, 1992.

Русская грамматика. Т. 2: Синтаксис. М., 1980.

Сусов И. П. Семантическая структура предложения. Тула, 1973.

Теньер Л. Основы структурного синтаксиса. М., 1988.

Филлмор Ч. 1) Дело о падеже // Новое в зарубежной лингвистике. Вып. X: Лингвистическая семантика. М., 1981а; 2) Дело о падеже открывается вновь // Новое в зарубежной лингвистике. Вып. X: Лингвистическая семантика. М., 1981б.

Хомский Н. Аспекты теории синтаксиса. М., 1972.

Чейф У. Значение и структура языка. М., 1975.

Шенк Р. Обработка концептуальной информации. М., 1980.

Якобсон Р. Шифтеры, глагольные категории и русский глагол // Принципы типологического анализа языков различного строя. М., 1972.

Cook W. Case Grammar: Development of the Matrix Model (1970—1978). Washington, 1979.

Zimek R. Sémantická vystavba věty. Praha, 1980.

МЕТОДЫ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ В ЯЗЫКОЗНАНИИ

ПОЗНАВАТЕЛЬНЫЕ ПРИНЦИПЫ СТАТИСТИЧЕСКОГО МЕТОДА

На фоне непреходящего интереса лингвистов к технике обработки данных в настоящее время на первый план выдвигается вопрос о месте и роли статистического метода в филологической науке и особенностях его прилаживания к решению конкретных задач. Не теряет актуальности и попытка разобраться в том, на что может претендовать статистический метод и на что он претендовать не может.

Статистический метод — это комплекс приемов и принципов, согласно которым производятся сбор, систематизация, обработка и интерпретация статистических данных с целью получения научных и практических выводов. Статистическими данными называют при этом количественные сведения о какой-либо совокупности объектов, обладающих теми или иными общими признаками, способными изменять свое качественное и количественное состояние. Математическое содержание комплекса приемов и принципов статистического метода образует математическая статистика — отрасль прикладной математики. В то же время статистический метод имеет и нематематическую сторону, к которой следует отнести логико-познавательные проблемы статистического исследования. Часть из них имеет универсальный характер, другая же часть определяется спецификой объекта и предмета каждой научной дисциплины.

Адаптация статистического метода к решению филологических проблем осуществляется в двух основных направлениях. Первое характеризуется стремлением к переосмыслению основных вероятностно-статистических категорий в филологических, преимущественно лингвистических, терминах. Вопрос здесь ставится так: что можно считать в филологии вероятностью, частотой, случайной величиной, совокупностью, выборкой, корреляцией, группировкой и т. п.? Для второго направления

характерно обратное движение мысли — стремление к вероятностно-статистическому переосмыслению основных лингвистических категорий, таких как «язык» и «речь», «норма» и «узус», «текст» и «стиль», «план выражения» и «план содержания», «поверхностный уровень» и «глубинный уровень», «парадигматика» и «синтагматика» и т. п. Между двумя тенденциями (филолого-статистической и статистико-филологической) нет четкой границы. Это, впрочем, не имеет принципиального значения, так как здесь важен сам факт органического слияния категорийно-понятийного аппарата языковедения и статистики. Наиболее полное представление о двойственных интерпретациях такого рода можно почерпнуть из работ В. В. Богданова, П. М. Алексеева и Ю. А. Тулдавы. К числу таких интерпретаций можно отнести и толкование текста как реальной совокупности (см. раздел «Стилеметрия»). Далеко не все среди предложенных исследователями интерпретаций представляются убедительными. Это относится в первую очередь к центральной вероятностно-статистической категории — «вероятность», которая в языковедении фигурирует как «лингвистическая вероятность» с различными вариантами: «языковая вероятность», «речевая вероятность», «парадигматическая вероятность», «синтагматическая вероятность», «нормирующая вероятность» и т. п.

Статистический метод, как уже было сказано, — универсальный метод познания действительности. Познавательные черты этого метода показаны на рис. 1. Здесь отражены взаимоотношения между субъектом и объектом исследования, объектом и методом, объектом и знанием о нем. Позиция исследователя характеризуется объективностью, беспристрастностью.

Интерес исследователя к объекту диктуется прежде всего практическими прикладными целями (например, задачей автоматического индексирования текста, его дешифровки или атрибуции), что на рис. 1 обозначено как прагматичность.

Взаимодействие исследователя с объектом осуществляется через статистический метод, характеризующийся одновременно и строгостью и приближенностью. Эти свойства не противоречат друг другу, так как в основе этого метода лежит теория вероятностей, все утверждения которой, будучи математически (формально) строгими, тем не менее имеют вероятностный, приближенный характер. Процедурность статистического метода подкрепляет позицию объективности исследователя. Приближенность метода тесно связана с познавательной установкой на прагматичность, так как при решении прикладных задач нет необходимости в исчерпывающем описании языка, здесь можно удовлетвориться приближенным описанием, достаточным для решения именно данной, а не какой-то другой задачи (Буторов, 1983).

Объект исследования представляет собой не жестко детер-

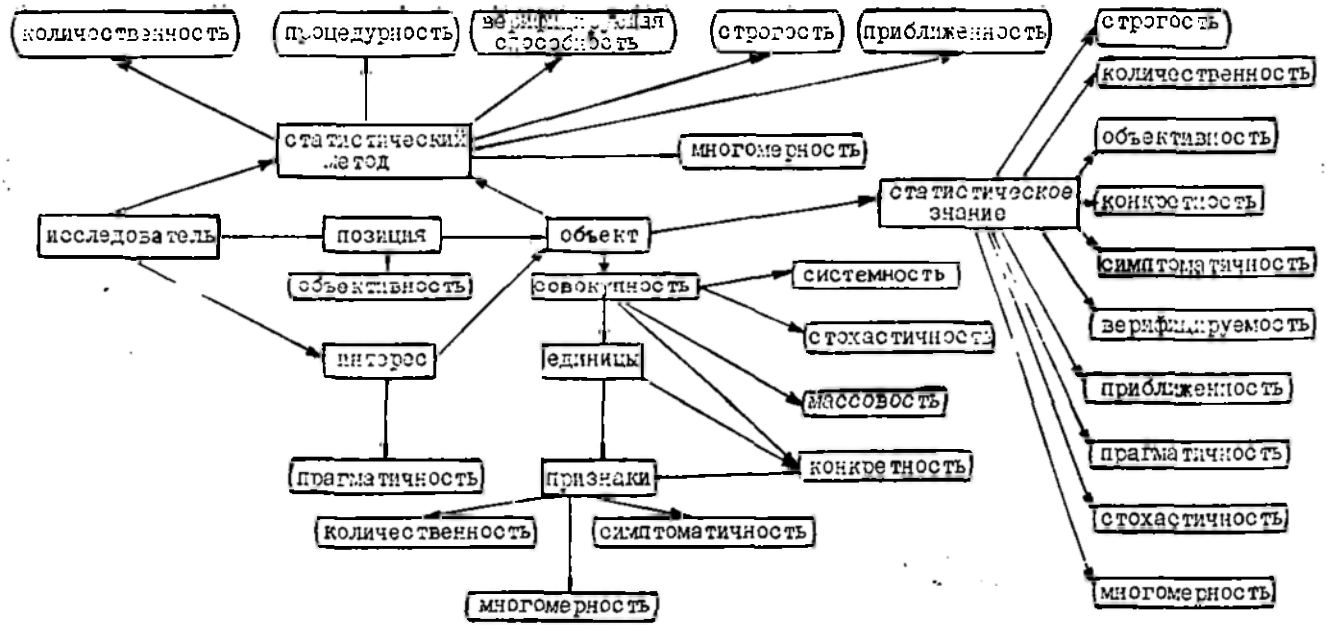

количественность
процедурность
строгость
приближенность
статистический метод
многомерность
исследователь
позиция
объект
объективность
совокупность
системность
стохастичность
интерес
единицы
массовость
прагматичность
признаки
конкретность
количественность
симптоматичность
многомерность
статистическое знание
строгость
количественность
объективность
конкретность
симптоматичность
верифицируемость
приближенность
прагматичность
стохастичность
многомерность

Рис. 1.

минированную, а мягкую, диффузную, вероятностную систему (стохастичность, системность), выступающую в виде совокупности (массовость) единиц, локализованных в конкретных рамках времени и пространства (конкретность) и обладающих множеством признаков (многомерность), среди которых исследователя интересуют не коренные, сущностные, а симптоматические признаки, имеющие количественную природу (количественность, симптоматичность)[1].

В результате взаимодействия исследователя (через интерес, позицию и метод) с объектом рождается статистическое знание, обладающее свойствами, вытекающими из позиции и интереса исследователя (объективность, прагматичность), а также объекта исследования (системность, стохастичность, симптоматичность, многомерность, конкретность, количественность) и метода исследования (строгость, приближенность, верифицируемость). Обращает на себя внимание тот факт, что метод взаимодействует не только с объектом, но и добытым знанием, отсюда следует свойство верифицируемости знания.

Познавательные принципы статистики во многом перекликаются с теми, которыми пользуется ортодоксальная дескриптивная лингвистика. Если учесть, что основным рабочим средством дескриптивной лингвистики является дистрибуция, а описательной статистики — распределение, то аналогия окажется еще более сильной. В этом контексте, например, в составном термине «дистрибутивно-статистический метод» оба компонента дублируют друг друга, ибо любой статистический метод является дистрибутивным. Столь же тавтологично (терминологически, не сущностно), но в другом контексте звучит термин «статистико-комбинаторный метод», ибо любая вероятностно-статистическая модель основана на комбинаторике.

ОСНОВНЫЕ МЕТОДЫ ОБРАБОТКИ ЛИНГВОСТАТИСТИЧЕСКИХ ДАННЫХ

В традиционной статистике различаются две группы методов: описательные методы и методы оценивания.

Задача первых заключается в том, чтобы представить исходные данные в компактной и наглядной форме (в виде таб-

[1] В. Г. Адмони противопоставляет симптоматическое, качественно-количественное описание, пользующееся простейшими подсчетами, статистическому методу, оснащенному изощренными средствами математической обработки лингвистических данных (Адмони, 1964). Л. Р. Зиндер справедливо отмечает, что такое противопоставление является неестественным, так как любая статистика является симптоматической (Зиндер, Строева, 1968). Что касается акцента, который делается на качественности симптоматического метода, то любое разумно организованное статистическое исследование требует качественного обоснования и качественной интерпретации: голые числовые изыскания без руководящей точки зрения в статистике не приветствуются.

лиц, графиков) и обобщить эти данные с помощью различного рода статистик: мер центральной тенденции, мер колеблемости, мер концентрации, мер связи и т. п.

К описательным методам тесно примыкает метод моделирования, заключающийся в выборе вероятностно-статистической модели, адекватно описывающей эмпирический материал.

В статистике долгое время доминирующее положение занимало нормальное (гауссово) распределение, затем на авансцену вышли другие распределения, так или иначе связанные с нормальным (пуассоновское, биномиальное, логарифмически нормальное и др.). В настоящее время позиции нормального закона в значительной степени подорваны. Так, при моделировании социальных процессов все большую силу набирает концепция негауссовости статистических распределений (Яблонский, 1975; Хайтун, 1973; Мартыненко, 1987). Отметим в этой связи смелые идеи Г. Хердана. Он считает, что статистика на фонологическом уровне и статистика на уровне словаря являются совершенно разными математическими дисциплинами. Фонемная статистика опирается на классический аппарат статистики и теории вероятностей, тогда как словарная статистика — на комбинаторные модели, заимствованные из аппарата статистической физики (Herdan, 1964).

Методы оценивания распадаются на две группы: методы оценивания неизвестных параметров распределения и методы проверки статистических гипотез. Первая группа методов решает вопрос о сходимости наблюдаемых величин к вероятностным. Она основана на использовании закона больших чисел и классических предельных теорем (теоремы Чебышева, центральной предельной теоремы Ляпунова). В центре классической теории оценивания стоит нормальное распределение, на основе которого строятся доверительные интервалы для параметров распределения. Использование тех или иных способов оценивания параметров распределения в значительной степени зависит от природы исследуемых совокупностей.

Если генеральная совокупность представляет собой разделительное множество бесконечного (или практически бесконечного) объема (о соотношении собирательных и разделительных категорий см. в разделе «Стилеметрия»), то ее исследование осуществляется путем формирования репрезентативной выборки, на основании которой вычисляются статистические оценки, сходящиеся к предельным величинам, рассматриваемым в теории вероятностей. При этом весьма актуальной является задача оценивания параметров не только нормально распределенной генеральной совокупности, но и любых других совокупностей. Особым случаем оценивания является использование экспресс-методов, когда для оценивания генеральной совокупности используются малые и очень малые выборки.

Но генеральная совокупность может быть представлена и

разделительным множеством конечного объема. Если она при этом не очень велика, то отпадает необходимость в выборочном обследовании. Средние величины в такой ситуации имеют точное значение и не нуждаются в теоретико-вероятностном оценивании.

Особого подхода требуют совокупности, представленные собирательными множествами (т. е. относящиеся к классу реальных совокупностей по А. А. Чупрову, см. раздел «Стилеметрия»). Такие совокупности могут быть сравнительно небольшими (и тогда их можно исследовать целиком), или же они могут содержать большое (но всегда конечное) число компонентов. В последнем случае встает вопрос о формировании представительной выборки. В ряде исследований показано, что для определенного класса таких совокупностей — тех, которые образуют собирательное множество, представительная выборка отсутствует (точнее, самая удачная из них будет того же порядка величины, что и генеральная совокупность). В этой ситуации важен не объем выборки, а факт целостности совокупности (например, текста). Некоторые исследователи: Ю. К. Орлов (1970), Ю. А. Шрейдер (1982), С. Д. Хайтун (1983) — полагают, что идея о сходимости традиционных характеристик (например, моментов) к предельным величинам здесь теряет смысл.

Что касается методов проверки статистических гипотез, то здесь в последние годы все большее распространение начинают получать непараметрические и устойчивые методы, основанные на весьма слабых допущениях о природе исследуемых совокупностей и законе их распределения. В качестве примера можно привести дисперсионный анализ: он абсолютно нечувствителен к отклонениям от нормальности, а в том случае, когда объемы сопоставляемых групп равны, он нечувствителен и к влиянию неоднородности дисперсий.

Более полное представление о методах обработки лингвостатистических данных можно получить путем анализа типов используемых в языкознании распределений.

ТИПОЛОГИЯ ЛИНГВОСТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Статистические ряды (распределения) представляют собой упорядоченную совокупность количественно выраженных значений, т. е. результатов измерения объекта (объектов), обычно с указанием статистического веса (частоты или ранга) этих значений в совокупности.

Теоретическая значимость статистических распределений определяется прежде всего тем, что естественные языки и тексты обладают рядом принципиальных свойств, присущих вероятностным системам: единством порядка и беспорядка в классе событий, единством иррегулярности и устойчивости, зависимости и автономности в классе событий. Что касается роли стати-

стических распределений в конкретных исследованиях, то они являются не только удобным, но и обязательным средством компактного представления, анализа, обобщения и интерпретации данных наблюдения и эксперимента.

Ниже приводится типология лингвостатистических распределений. При определении степени существенности «типообразующих» признаков мы руководствовались, в первую очередь, теми сторонами лингвистической реальности, которые могут быть интерпретированы в системном плане, а лишь затем — формально-математическими соображениями.

Предлагаемая система типов распределения не претендует на исчерпывающую полноту, это, скорее, перечень аспектов, построенных в виде бинарных противопоставлений, в рамках которых выделяются конкретные типы статистических рядов.

1. **Вероятностная теория — лингвистическая реальность.** В рамках этого противопоставления можно говорить об эмпирических рядах, отражающих результат группировки данных наблюдений, и теоретических распределениях, выбранных для описания закономерности, которой подчиняется фактическое распределение.

Хотя число потенциалов возможных теоретических распределений очень велико, все многообразие эмпирических рядов стараются свести к хорошо исследованным теоретическим вариантам. Многие из них занимают в статистике особое положение, либо потому, что они обладают желательными математическими свойствами, либо потому, что описываемый ими механизм вариации строго или хотя бы приблизительно соответствует вариации, характерной для конкретного фрагмента действительности. Важной чертой теоретических распределений является то, что многие из них могут рассматриваться как частные случаи более широких семейств.

Среди принципов, на основании которых выделяются конкретные семейства распределений, можно назвать следующие.

Распределения могут различаться с точки зрения степени привлечения теоретико-вероятностных комбинаторных схем для объяснения причинных факторов, регулирующих распределение вероятностей. Некоторые распределения (нормальное, биномиальное, пуассоновское, паскалево и др.) имеют строгое теоретико-вероятностное обоснование, другие такого обоснования не имеют (например, распределение Вейбулла) или имеют частичное обоснование (например, распределение Ципфа).

То или иное семейство распределений может строиться путем трансформации эталонного распределения (например, нормального) в функционально с ним связанное. Так строится, например, семейство распределений Кэмптейна.

И наконец, семейство распределений может строиться с использованием свертки (композиции) нескольких распределений одного вида и даже сильно различающихся распределений. На-

пример, композиция нескольких распределений равномерной плотности близка к нормальному распределению.

Теоретические распределения четко распадаются на две группы: моделирующие распределения и критериальные распределения. Первые используются для аппроксимации эмпирических распределений, вторые — для оценивания неизвестных параметров и проверки статистических гипотез (распределение Стьюдента, Пирсона, Фишера и др.).

2. **Гауссовость — негауссовость.** С теоретико-вероятностной точки зрения все реальные и мыслимые варианты распределений распадаются на две группы: семейство гауссовых распределений и семейство негауссовых распределений. К первой группе относятся теоретические распределения, описываемые тем или иным вариантом экспоненциальной (показательной) функции. В роли эталона этой группы выступает нормальное распределение, к которому сходятся все варианты гауссового семейства кривых распределений (биномиальное распределение, распределение Пуассона, распределение Стьюдента и многие другие). В роли эталона второй группы выступает уравнение неравносторонней гиперболы. Это уравнение считается идеальным образцом (или асимптотикой), к которому устремляются все варианты распределений гиперболического типа (распределения Ципфа — Парето, Эсту — Лотки, Мандельброта и др.). Отличительной чертой негауссовых распределений является бесконечность моментов любого порядка, в частности дисперсии.

История изучения негауссовых распределений сравнительно молода. Их основные свойства были впервые исследованы П. Леви, который, в частности, установил, что при композиции двух распределений гиперболического типа получается распределение того же вида. Это означает, что негауссовы распределения с теоретико-вероятностной точки зрения принадлежат к классу устойчивых распределений (Lévy, 1925).

3. **Типичность — нетипичность.** Делению теоретических распределений на гауссовы и негауссовы в традиционной статистике соответствует деление эмпирических рядов на типические и нетипические.

С содержательной точки зрения такое деление в статистике связывается с делением всех явлений, к какой бы области они ни принадлежали, на две группы: явления типические и явления индивидуальные. Первые — это те, где доминируют постоянные причины, вторые — где решающее влияние принадлежит случайным причинам. С математической точки зрения типическими рядами считаются те, в которых механизм взаимодействия постоянных и случайных причин строго или хотя бы приблизительно соответствует моделям массовых процессов, обобщенных в первоначальных теоремах закона больших чисел. Все остальные ряды рассматриваются как нетипические.

4. **Ранг — частота.** Каждое теоретическое распределение имеет вид функции, устанавливающей связь между возможными значениями случайной величины и соответствующими им вероятностями. Существует несколько способов задания этой связи. Важнейшие из них представлены в табл. 1, а в табл. 2 при-

Таблица 1. **Способы задания функциональной связи между случайной величиной и вероятностью**

Теоретико-*вероятностный* термин	*Описание*	Математико-*статистический* термин	Лингвостатистический термин
Функция распределения $F(x)$	$F(x)$ есть вероятность того, что случайная величина принимает значение, меньшее или равное x: $F(x) = \text{Prob}\,[X < x] = \alpha$, $F(x) = \int_{-\infty}^{x} f(u)\,du$	Кумулятивный ряд	—
Плотность вероятности $f(x)$	Функция, интеграл от которой по промежутку от x_L до x_u равен вероятности того, что случайная величина принимает значение из этого промежутка: $\int_{x_L}^{x_u} f(x)\,dx = \text{Prob}\,[x_L X < x_u]$, $f(x) = \frac{d[F(x)]}{dx}$	Вариационный ряд	Спектровое распределение
Обратная функция распределения (функция квантилей) $G(\alpha)$	$G(\alpha)$ — такое число, что случайная величина примет значение, не превосходящее α, с вероятностью $G(x)$: $x = G(\alpha) = G[F(x)]$, $\text{Prob}\,[X < G(\alpha)] = \alpha$	Возрастающий ранжированный ряд	—
Функция выживания* $S(x)$	$S(x)$ — это вероятность того, что случайная величина примет значение большее, чем x: $S(x) = \text{Prob}\,[X > x] = 1 - F(x)$	—	—
Обратная функция выживания (обратная функция квантилей) $Z(\alpha)$	$Z(\alpha)$ — это такое значение, которое случайная величина превосходит с вероятностью α: $\text{Prob}\,[X > Z(\alpha)] = \alpha$, $x = Z(\alpha) = Z[S(x)]$, где $S(x)$ — функция выживания, $Z(\alpha) = G(1 - \alpha)$, где G — обратная функция распределения	Убывающий ранжированный ряд	Ранговое распределение

* Термин заимствован из демографических таблиц смертности, где используются «вероятности выживания» — вероятности дожить до определенного возраста.

ведены варианты двух распределений простейшего типа: распределения Ципфа — Парето и показательного распределения. Обе таблицы построены на основании сведений, систематизированных Н. Хастингсом и Дж. Пикоком (1980).

Таблица 2. **Формы функционального задания показательного распределения и распределения Ципфа — Парето**

Форма задания распределения	Распределение Ципфа—Парето	Показательное распределение	Обозначения и примечания
Кумулятивный ряд	$K - Kx^{-c}$	$K - Ke^{-\lambda x}$	K — объем совокупности, x — величина варьирующего признака, r — ранг
Вариационный ряд	$Kcx^{-(c+1)}$	$K\lambda e^{-\lambda x}$	
Возрастающий ранжированный ряд	$\left(\frac{K}{K-r}\right)^{\frac{1}{c}}$	$\frac{1}{\lambda}\ln\frac{K}{K-r}$	
Декумулятивный ряд	Kx^{-c}	$K\lambda e^{-\lambda x}$	
Убывающий ранжированный ряд	$\left(\frac{K}{r}\right)^{\frac{1}{c}}$	$\frac{1}{\lambda}\ln\frac{K}{r}$	

Из табл. 1 видно, что разные способы задания распределения находятся в отношениях взаимно-однозначных отображений: интегральная форма — дифференциальная, прямая функция — обратная функция, вероятность события — вероятностность противоположного события. В лингвостатистических исследованиях связь между различными формами задания одного и того же распределения часто не осознается, в частности, неучитывается связь между двумя формами рангового распределения (возрастающим и убывающим), а также между функцией распределения и ранговым распределением.

В табл. 2 показано, что между спектровым распределением и убывающим ранговым распределением нет никакой сущностной разницы, это всего лишь разные формы задания одного и того же распределения. Основные геометрические типы распределений представлены на рис. 2 в графе «Распределения простейшего типа».

Кроме упомянутых распределений можно было бы выделить еще два кумулятивных распределения: для возрастающего и убывающего ранговых распределений.

5. **Разнообразие — ограничение разнообразия.** Порождение текста протекает на фоне двух тенденций: стремления к расширению состава разноименных единиц (увеличение разнообразия) и стремления к увеличению частоты отдельных единиц (ограничение разнообразия). Этим противоборствующим тенденциям соответствует два типа количественной вариативности лингвистических единиц: вариативность состава (номенклатурная колеблемость знакотипов) и рекуррентная вариативность.

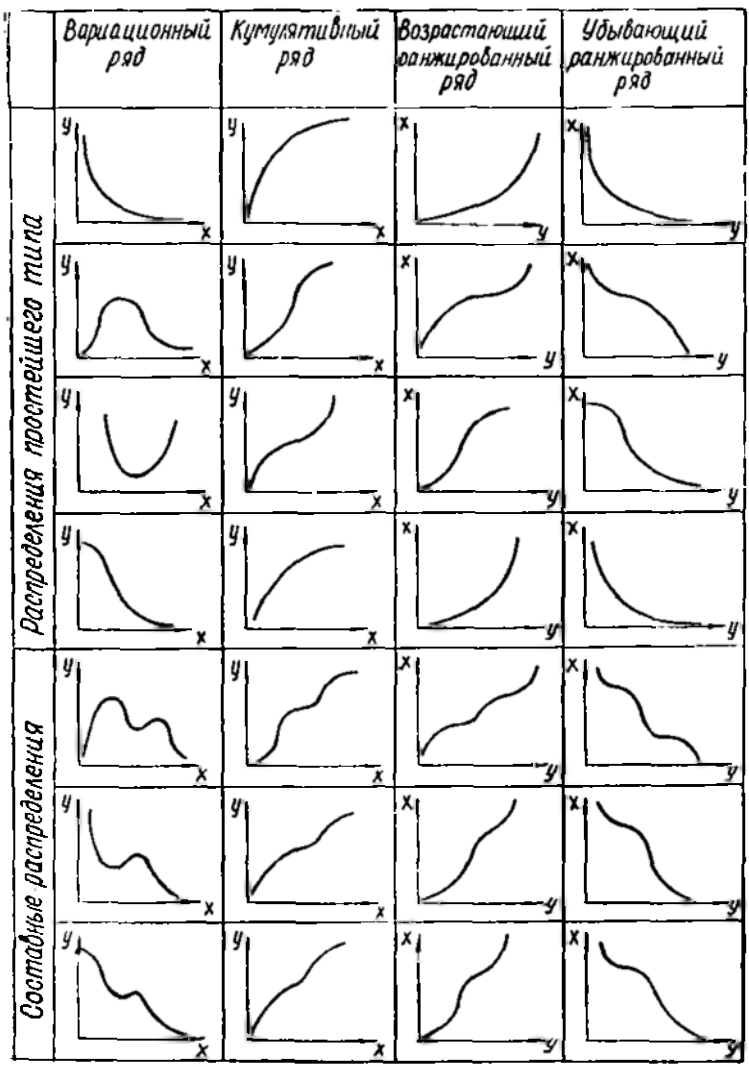

Вариационный ряд
Кумулятивный ряд
Возрастающий ранжированный ряд
Убывающий ранжированный ряд
Распределения простейшего типа
Составные распределения

Рис. 2.

(колеблемость числа знакоупотреблений каждого знакотипа и знакотипов в целом). Распределения, связывающие накопляемое количество знакотипов с накопляемым числом знакоупотреблений, принято называть распределением знакотип — знакоупотребление. В статистической лексикографии этот тип распределений используется при вычислении индексов лексического разнообразия (богатства) и лексической концентрации (покрываемости) текстов.

При построении распределения знакотипов отвлекаются от рекуррентной вариативности единиц текста, при построении распределения знакоупотреблений учитываются оба вида вариативности. Иногда эти распределения называют соответственно распределениями в словаре и распределениями в тексте.

6. **Строение — поведение.** Объект познания — то, на что направлена познавательная деятельность, представляет единство двух противоположных сторон: поведения и его материального носителя — строения. Из этого следует, что при построении статистических рядов могут быть использованы две категории варьирующих признаков: 1) признаки, отражающие внутреннюю структуру единиц совокупности; 2) признаки, указывающие на статус (функциональный вес) единиц, образующих совокупность.

В соответствии с этими категориями признаков можно говорить о двух типах статистических рядов: о статусных рядах и структурных рядах. Текстовые словоформы могут быть упорядочены, например, по их размеру (числу букв, фонем, морфем), т. е. по признаку, отражающему внутреннюю структуру словоформы. В этом случае «коллектив» словоформ образует структурное распределение. Но та же совокупность словоформ может быть проранжирована по функциональному признаку, например, по их встречаемости в тексте. В этом случае образуется статусное распределение. Между двумя типами распределений нет абсолютно четкой границы: функции конкретного объекта определяются тем, как он устроен, и, наоборот, внутреннее строение объекта в значительной мере определяется его функциями. Связь между статусными и структурными распределениями характерна для организованных коллективов любой природы: от неорганической до знаково-информационной. Например, весовое содержание конкретного химического элемента в различных средах зависит от его атомного веса, численность популяции в биоценозе — от размера особи, а частота слова в тексте — от его размера, степени полисемии, объема понятия и т. п.

7. **Элемент — совокупность элементов.** Статистическое описание совокупности объектов занимает промежуточное положение между индивидуальным описанием каждого из объектов совокупности в отдельности и описанием совокупности по ее общим

свойствам, совсем не требующим ее расчленения на отдельные объекты. Статистический интерес может смещаться или в сторону изучения свойств отдельного объекта, или в сторону изучения совокупности объектов. В первом случае лингвиста интересует поведение одного и того же объекта в различных ситуациях (однопредметное наблюдение). Во втором случае исследуется поведение различных, но родственных в каком-либо отношении единиц в одной и той же группе или группе однородных ситуаций (многопредметное наблюдение). Различие в целях исследования ведет к построению двух контрастных типов распределений. Распределения первого типа назовем однопредметными, а распределения второго типа — многопредметными. Принципиальная непохожесть этих распределений была впервые осознана основоположником научной статистики А. Кетле, который заметил резкую грань, отделяющую средние величины, полученные при многократном измерении признака одного и того же объекта (например, высоты одного и того же дома) и однократном измерении величины разнокачественных, но родственных в каком-либо отношении объектов (например, высоты разных домов на одной и той же улице).

Различие между однопредметными и многопредметными распределениями с особой откровенностью проявляется тогда, когда результаты эксперимента представлены в виде двухмерной таблицы. Примером такой таблицы могут служить результаты анализа цветописи Гоголя, осуществленного А. Белым (1934).

Данные, приведенные в строках табл. 3, могут рассматриваться как однопредметные распределения (вариация конкретного цвета по группам произведений), а данные столбцов — как многопредметные распределения (вариации цветовой гаммы каждой группы произведений). Эта пространственная ориентация нашла отражение в терминологии, уже получившей некоторое распространение в статистической лексикографии: однопредметные распределения называют горизонтальными, а многопредметные — вертикальными распределениями.

Отметим, что в роли «предметов» конкретного распределения могут выступать объекты разной степени общности: от самых элементарных, например, тех, которые используются в обычных частотных словарях, до предельно обобщенных, например, тех, которые сконструировал Андрей Белый. А это означает, что результаты статистического эксперимента находятся под сильным влиянием внестатистической концепции, положенной в основу формирования обобщенных объектов, а также интуиции, изобретательности и даже смелости исследователя.

8. **Виртуальность — актуальность.** В лингвостатистике различают два типа совокупностей в зависимости от того, к какому уровню познавательной деятельности они относятся: к уровню

Таблица 3. **Цветовой спектр Гоголя (по А. Белому)**

Цвет	Группа произведений				Статистическая средняя, %
	I	II	III	IV	
Красное	26,6	12,5	10,3	6,4	17,4
Белое	9,5	9,0	22,0	17,0	14,0
Черное	11,0	14,1	11,8	4,8	12,0
Зеленое	8,6	7,7	9,6	21,6	9,4
Золотое	11,6	8,9	2,8	12,8	9,2
Синее	10,7	6,1	4,9	6,4	8,7
Желтое	3,5	8,5	10,3	12,8	7,0
Серое	2,6	8,9	10,5	6,4	5,8
Голубое	4,4	5,7	7,0	1,6	4,8
Серебряное	7,1	3,2	2,8	4,8	4,8
Коричневое	0,9	6,5	8,4	1,6	4,0
Розовое	3,8	0,8	2,1	3,2	2,3
Оранжевое	0,3	1,6	2,8	0,0	1,2
Лиловое	1,7	1,2	0,0	1,6	0,9

Обозначения: I — «Вечера на хуторе близ Диканьки», «Вий» и «Тарас Бульба»; II — повести и комедии; III — первый том «Мертвых душ»; IV — второй том «Мертвых душ».

непосредственного наблюдения или к сфере объективизации. В первом случае исследуется вариативность лингвистических единиц в тексте, во втором — в продуктах упорядочивающей, систематизирующей и обобщающей деятельности лингвиста: терминологических и толковых словарях, семантических сетях, информационно-поисковых тезаурусах и т. п. Распределения, строящиеся на материале объективизированных (виртуальных) единиц, называют распределениями в сфере фиксации, а распределения, строящиеся на материале текстовых (актуальных) единиц, — распределениями в сфере функционирования.

9. **Однородность — неоднородность.** С точки зрения классических математико-статистических представлений основным условием, определяющим совокупность в статистическом смысле, является однородность ее единиц по определенному признаку. При этом под однородностью понимается не только качественная, но и количественная характеристика. Однородность совокупности, определяемую на основе количественного признака, называют количественной однородностью, а на основе качественного признака — качественной однородностью.

В качестве условия однородности в статистике используется критерий близости значений количественного признака, который при этом выступает не самостоятельно, а в необходимом единстве с условием однокачественности явлений. Количественная однородность только тогда имеет смысл, когда отража-

ет единство качества и количества, т. е. является мерой такого единства. Отклонения от нее свидетельствуют о переходе одного качества в другое на основе накопленных качественных и количественных изменений.

Основным критерием, позволяющим судить, является ли совокупность однородной в количественном отношении, по сей день является критерий нормальности. Этот критерий еще в 20-х годах предложил известный советский статистик Б. С. Ястремский. Он, в частности, писал следующее: «Варианты признака однородной совокупности обладают общим уровнем, отклонения от которого оказываются случайными: ни один из множества факторов не выделяется своим влиянием на общий результат. Именно такой характер влечет за собой в ряде случаев образование нормального распределения, аналогично закону распределения погрешностей.

Варианты признака разнородной совокупности лишены общего уровня и вследствие этого не укладываются в нормальное распределение. Наличие асимметричного распределения сигнализирует о необходимости тщательных поисков возможных источников разнородности» (Ястремский, 1960, с. 92).

Развивая подход, предложенный Б. С. Ястремским, В. И. Сиськов разработал алгоритм построения однородных в количественном отношении групп. В качестве критерия однородности В. И. Сиськов предлагает использовать коэффициент вариации: группа признается однородной, если значение коэффициента вариации в ней не превышает 0,33 (Сиськов, 1975).

В большинстве случаев вопрос формирования однородной совокупности решается на стадии планирования наблюдения путем качественного анализа с последующим расчленением неоднородной совокупности на качественно однородные группы. В результате такого анализа строится несколько распределений, состоящих из однородных в качественном отношении единиц. Однако в ряде случаев качественная специфика лингвистических объектов такова, что прямого распознавания принадлежности конкретного объекта к данной подсовокупности осуществить не удается, т. е. неоднородность априорно не устранима. Это дает основание говорить о распределениях с априорно устранимой и распределениях с априорно не устранимой неоднородностью. Последние возникают в том случае, когда между подмножествами лингвистических единиц нет четкой границы, она может быть установлена лишь условно с помощью формальных процедур, разрабатываемых в теории статистических группировок.

Иногда однородность — неоднородность трактуется не с точки зрения сходства единиц, образующих группы, а с точки зрения близости распределений в разных группах. В этом случае с помощью критериев согласия по стандартной схеме проверя-

ется принадлежность двух выборок (групп) к одной и той же генеральной совокупности.

10. **Устойчивость — неустойчивость.** Проблема устойчивости статистических чисел является центром, вокруг которого уже не одно десятилетие вращается статистическая мысль. Однако длительность общения с этой проблемой не привела к серьезным сдвигам в ее решении. Даже сегодня некоторые ученые рискуют утверждать, что «вопрос о причинах устойчивости частот можно считать только поставленным. Его решение еще не предвидится, хотя некоторые усилия в этом направлении предпринимаются» (Сачков, 1971, с. 123).

Трудность решения этой проблемы заключается в том, что в явлениях реальной действительности нет той абсолютной устойчивости, которая характерна для схем, конструируемых теорией вероятностей. В статистике, включая лингвистическую, принято говорить об относительной устойчивости статистических рядов, сравнительно медленно изменяющихся во времени и пространстве. Это означает, что при интерпретации эмпирической реальности в вероятностных терминах нужно соблюдать известную осторожность. Ведь считается чуть ли не аксиомой, что аппарат теории вероятностей может быть применен лишь к таким явлениям, где частоты обладают устойчивостью. И все же эта аксиома не останавливает статистиков-практиков в их усилиях познать закономерности реальных массовых процессов именно на почве теории вероятностей. Эти усилия не лишены оснований.

Во-первых, роль высшего арбитра, «примиряющего» теорию с практикой, выполняет закон больших чисел, содержание которого не исчерпывается содержанием математическим. Это закон объективной действительности, внешне проявляющийся в устойчивости статистических чисел. «Устойчивость статистических чисел, их свойство колебаться из года к году, лишь в известных, ограниченных пределах, — писал А. А. Чупров, — представляет собой эмпирически установленный факт, который сам по себе, независимо от тех или иных теоретических истолкований, имеет громадную научную и жизненную важность. Это один из коренных, хоть и мало заметных, устоев современной культуры» (Чупров, 1910).

Во-вторых, хотя все модели, конструируемые в теории вероятностей, начиная с теоремы Бернулли, ориентированы на испытания с регламентированной, целесообразно организованной случайностью, условия, при которых предоставляется роль случаю, в них постоянно расширяются и усложняются, все более приближаясь к условиям взаимодействия случайности и необходимости в явлениях реальной действительности.

В-третьих, любой разумно организованный статистический эксперимент может обладать практической (и не только практической) ценностью и без обращения к теоретико-вероятност-

ным схемам, но «лишь на почве теории вероятностей удается осмыслить наблюдаемые в статистических числах колебания, лишь теория вероятностей открывает возможность прагматического их истолкования в каждом отдельном случае» (Чупров, 1910).

Перечисленные факторы в совокупности создают основу для разыскания критериев устойчивости эмпирических распределений.

Первый шаг в этом направлении был сделан немецким статистиком В. Лексисом (Lexis, 1903), предложившим для измерения устойчивости статистических рядов особый показатель Q^2 — коэффициент дивергенции. Смысл этого показателя заключается в том, что эмпирическая дисперсия сравнивается с теоретической величиной, условно определяемой таким образом, что частость, взятая для всего ряда данных, приравнивается вероятности:

$$Q^2 = \frac{2\sum(p_i - p)^2}{m-1} : \frac{2p(1-p)}{n},$$

где m — число серий по n испытаний, p_i — частота в сериях, p — общая частость, приравненная вероятности.

По Лексису, небольшое отличие Q^2 от единицы свидетельствует о нормальном рассеянии, если Q^2 меньше единицы, то устойчивость ряда сверхнормальна, а если больше единицы, то устойчивость ряда поднормальна.

В соответствии с величиной Q^2 можно говорить о рядах с нормальной, поднормальной и сверхнормальной устойчивостью.

В длинной серии статистических рядов, исследованных Лексисом, ни в одном случае не получилось сверхнормальной устойчивости, иначе говоря, та степень устойчивости, которая характерна для закона ошибок Гаусса, имеет, по Лексису, значение максимальной. В подавляющем большинстве случаев устойчивость эмпирических рядов оказалась ниже нормы, зачастую уступая последней в 5, 10 и даже в 20 раз.

Экспериментальные результаты, полученные Лексисом и его многочисленными последователями, были в дальнейшем пересмотрены А. А. Чупровым, которому удалось показать, что в явлениях действительной жизни ряды со сверхнормальной устойчивостью встречаются достаточно часто. Примечательно, что этот факт был им обнаружен на лингвистическом материале — в распределении букв в сериях выборок из произведений Гёте, в распределении дактилей в гекзаметрах Овидия и Вергилия.

В настоящее время в качестве мер устойчивости используются самые разнообразные показатели. Чаще всего применяется критерий χ^2, иногда с неоправданной «лингвистической коррекцией». К числу таких коррекций относится толкование

результатов применения критерия χ^2 Б. Н. Головиным: «...Полезно было бы различать три степени статистической устойчивости явлений языка и речи: первая (оптимальная устойчивость) — показатель „хи-квадрат" не выходит за пределы границы существенности, вторая (нарушаемая устойчивость) — величина „хи-квадрат" превышает границы существенности, но не более чем в два раза, третья (неустойчивость) — величина „хи-квадрат" превосходит границы существенности более чем в два раза» (Головин, 1971, с. 90—91).

Реже используется традиционный коэффициент вариации, а также его модификации (наиболее известной из них является коэффициент стабильности Жийяна) и совсем редко — показатели, основанные на учете мер частотности и распространенности. К их числу относятся коэффициент специфичности, основанный на учете распространенности элемента (представленности в микровыборках) (Андреев, 1965), и коэффициенты лексической устойчивости, учитывающие и фактор частоты, и фактор распространенности (Кобрин, 1976; Марусенко, 1980).

11. **Редкость — частость.** В классической статистике работа с частыми элементами осуществляется с помощью нормального закона и предельных теорем, основу которых он образует. Что касается редких событий, то здесь используется закон Пуассона (закон редких событий) как при моделировании реальных процессов, так и при оценивании неизвестных параметров распределения.

В языкознании различие между распределениями редких и частых элементов трактуется следующим образом: частые (семантически стертые, недоминантные) единицы текста подчиняются нормальному закону, а редкие (семантически ключевые, доминантные) единицы — закону Пуассона. Если речь идет, например, о художественном произведении, то семантически доминантными будут слова и словосочетания, обладающие с точки зрения сюжета и композиции важным значением. Р. Г. Пиотровский объясняет этот факт тем, что «порождение текста определяется, с одной стороны, системой языка и его нормой, а с другой — независимой от языка ситуацией, которую призван описывать этот текст. Появление семантически нагруженных ключевых элементов текста диктуется ситуацией, употребление же семантически „стертых" единиц, так называемых единиц заполнения, служащих средством организации текста и связи ключевых единиц, подчиняется требованиям системы и нормы языка» (Пиотровский, Бектаев, Пиотровская, 1977, с. 356—357).

Употребление частых и редких элементов в тексте обсуждается и с точки зрения их роли в целостной системе, а также их приоритетности с точки зрения исследовательского интереса (Арапов, Шрейдер, 1970). Этот вопрос имеет большое значение для стилеметрических задач, в частности, для выяснения роли

редких и частых элементов при атрибуции и типологии текстов (Орлов, 1970; Борода, 1977).

12. **Симметрия — асимметрия.** Встречи с симметричным распределением в лингвистике сравнительно редки. В подавляющем большинстве случаев исследователь имеет дело с умеренно и крайне асимметричным распределениями (см. рис. 2). Та или иная степень асимметрии может возникать вследствие: 1) соединения в одной совокупности качественно неоднородных элементов или группы элементов; 2) наличия связей между единицами совокупности; 3) ограниченности начальных или крайних значений варьирующего признака каким-либо числом; 4) принадлежности лингвистической единицы к конкретному лингвистическому уровню или лексикограмматическому классу, а также по другим причинам, с трудом поддающимся систематическому учету.

13. **Одновершинность — многовершинность.** Наряду с одновершинными (одномодальными) в лингвостатистических исследованиях часто возникают многовершинные (полимодальные) вариационные ряды (см. рис. 2). Появление многовершинности ряда почти всегда указывает на неоднородность состава того коллектива, который дал материал для построения ряда. При этом распределению некоторых явлений свойственна полимодальность по существу особенностей их строения. Впервые на одно из таких лингвистических явлений обратил внимание известный русский статистик П. А. Вихляев. На материале произведений Тацита, Цезаря, Данте и Леопарди он показал, что распределения словоупотреблений по размеру всегда двухвершинны. Причина этого явления, по мнению Вихляева, «кроется в самом свойстве языка, который слагается из коротких частей — предлогов и союзов и более длинных — прилагательных и существительных» (Вихляев, 1928).

Многовершинность распределений часто связывается с явлением трансгрессии, заключающимся в том, что при «слиянии» двух совокупностей максимум распределений значений признака одной из них заходит за минимум другой. Если между средними арифметическими таких рядов различия статистически достоверны, то такие ряды называются трансгрессирующими. Основные геометрические типы двухвершинных рядов приведены на рис. 3, 4, 5: на рис. 3 — распределение словоупотреблений по длине (по данным П. А. Вихляева, 1928) (d — размер словоупотребления, n — частотность); на рис. 4 — распределение неопределенного артикля среди детерминативов у персонажей шекспировских комедий (по данным А. Я. Шайкевича, 1982) (f — частота артикля, n — число персонажей); на рис. 5 — распределение числа различных предметов и явлений, встречающихся группами в тексте американской «Энциклопедии мелочей» (Worth, 1974) (n — объем группы, N — число групп данного объема).

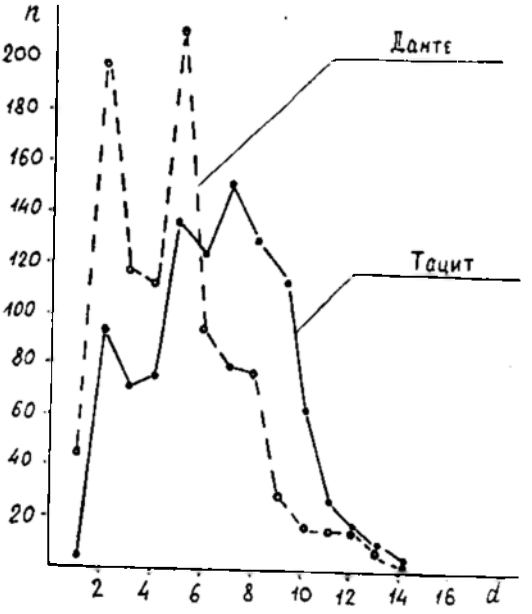


Рис. 3.

Приведем некоторые пояснения к рис. 4 и 5. Различие двух групп персонажей (см. рис. 4) связано как с социальными, так и со стилистическими факторами. Среди персонажей, у которых

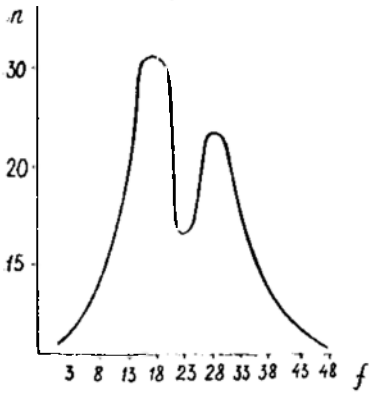


Рис. 4.

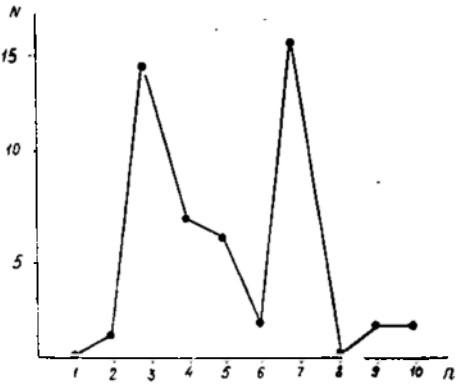


Рис. 5.

отмечено повышение частоты неопределенного артикля, преобладают слуги, горожане, а также те, кого можно назвать носителями комического начала пьесы (остряки или, напротив,

жертвы шуток, глупцы и т. п.). Наоборот, «серьезные люди», например монархи, характеризуются понижением доли неопределенного артикля в их речи. Различие двух групп чисел (см. рис. 5) определяется прежде всего тем, что в нашей жизни чаще всего употребляются «магические» числа «три» и «семь». Вспомним хотя бы о трех грациях (Аглае, Талии, Эвфросине — дочерях Зевса и Эвриномы), о трех картах (тройке, семерке, тузе), трех мушкетерах (Атосе, Портосе, Арамисе). Что касается числа семь, то и здесь очень много примеров: семь чудес света, семь цветов радуги, семь нянек и т. д.

Полимодальность лингвостатистических рядов иногда имеет скрытый характер. Это обычно наблюдается в распределениях с большим размахом вариации, производящих впечатление крайне асимметричных. Некоторые не слишком откровенные нерегулярности в поведении таких распределений обычно списываются на счет ограниченности объема выборки.

14. **Статика—динамика.** Изучение изменения явлений во времени является одной из важнейших задач статистики. Эта задача решается при помощи построения и анализа рядов динамики (динамических, или временных, рядов). Динамический ряд представляет собой ряд числовых значений конкретного статистического показателя в последовательные моменты или периоды времени, называемые уровнями ряда. В отличие от статических рядов, в которых в роли аргумента выступают признаки, характеризующие внутреннюю или внешнюю структуру единиц совокупности, в динамических рядах в роли аргумента выступают временные характеристики, формируемые на основе различных интерпретаций категории времени.

В зависимости от того, какой смысл вкладывается в категорию «время», различаются несколько разновидностей динамических рядов, применяемых в филологических исследованиях. Основное различие проходит в двух плоскостях: статика — динамика, синхрония — диахрония.

Первая разновидность связана с интерпретацией речепроизводства (текстообразование) как динамического процесса (динамика в отличие от статики). При этом динамика может трактоваться как «живой» порождающий процесс (с точки зрения статистики как нестационарный случайный процесс) или рассматриваться как застывший процесс, основные этапы которого регистрируются и анализируются исследователем с целью выявления какой-то тенденции (тренда). В последнем случае в центре внимания исследователя оказывается линейно-временная развертка текста. Текст мыслится при этом как последовательность присоединяемых друг к другу отрезков, в каждом из которых фиксируется конкретное значение случайной переменной. Но структура текста может пониматься и как последовательность сюжетных, композиционных, тематических, «персонажных» и т. п. частей, т. е. в данном случае используется худо-

жественное время. Такие ряды строил А. Белый, связывая ритмическую структуру произведения с сюжетно-тематическими моментами. Однако независимо от того, на какую интерпретацию текста мы опираемся (лингвистическую или литературоведческую), мы всегда пользуемся внутритекстовым временем. Ряды, которые строятся в такой ситуации, мы называем линейно-временными.

Вторым признаком, на основании которого осуществляется различение динамических рядов, как было сказано выше, является противопоставление синхронии и диахронии. Диахронические ряды, в отличие от синхронических, учитывают исторический, эволюционный фактор. В стилеметрическом аспекте можно изучить эволюцию индивидуально-стилевых, жанровых, родовых и других литературных систем. В этом случае время может пониматься в историческом, эпохальном, календарном и других смыслах.

Подведем некоторые итоги. Признаки, на основании которых выделены типы распределений, относятся к самым разным сторонам лингвистической и статистической реальности. Часть из них опирается на теоретико-вероятностные представления, другая — на способы задания и представления эмпирических данных, третья исходит из специфики исследуемых совокупностей, ее единиц и варьирующего признака. Не все из выделенных типов распределений обладают равнозначным «системным весом», однако каждый из них может оказаться в центре исследовательского интереса в зависимости от целей и задач наблюдения или эксперимента. Следует также иметь в виду, что избранный нами способ систематизации не устраняет частичной «синонимии» типов распределений, обусловленной в основном причинно-следственными отношениями между бинарными противопоставлениями. Так, асимметрия распределения и тем более его многовершинность могут рассматриваться как внешнее (геометрическое) проявление того или иного варианта качественной неоднородности, а степень устойчивости — как количественное выражение неоднородности.

В заключение отметим, что теория лингвостатистических распределений находится в стадии становления. Усилия лингвистов направлены в основном на то, чтобы добиться корректного и гармоничного сочетания понятийно-категориального аппарата статистической науки и науки о языке.

Литература

Адмони В. Г. Качественный и количественный анализ грамматических явлений // Теоретические проблемы современного советского языкознания. М., 1964. С. 56—68.

Алексеев П. М. Квантитативные аспекты речевой деятельности // Языковая норма и статистика. М., 1977. С. 43—58.

Алексеев П. М., Манасян Н. С. Лингвистические распределения (элементы количественного анализа текста). Л., 1985.

Андреев Н. Д. Вероятностные дифференциальные признаки и аппроксимационный анализ синтаксиса // Статистико-комбинаторное моделирование языков. М.; Л., 1965. С. 340—346.

Арапов М. В. Квантитативная лингвистика. М., 1978.

Белый Андрей. 1) Мастерство Гоголя. М.; Л., 1934; 2) Ритм как диалектика и «Медный всадник». М., 1929.

Богданов В. В. Статистические концепции языка и речи // Статистика речи и автоматический анализ текста 1972. Л., 1973. С. 9—19.

Борода М. Г. О характере распределения информационных единиц малой частоты в художественных текстах // Семиотика и информатика. Вып. 9. М., 1977.

Буторов В. Д. Приближенные методы анализа языка в прикладной лингвистике // Структурная и прикладная лингвистика. Вып. 2. Л., 1983. С. 131—137.

Вихляев П. А. Очерк теоретической статистики. М., 1928.

Головин Б. Н. Язык и статистика. М., 1971.

Зиндер Л. Р., Строева Т. В. К вопросу о применении статистики в языкознании // Вопр. языкозн. 1968. № 6. С. 120—123.

Кобрин Р. Ю. Лингвостатический анализ терминологических систем // Вычислительная лингвистика. М., 1976. С. 144—158.

Марусенко М. А. Коэффициент статистической устойчивости как критерий отбора учебного материала // Материалы семинара «Статистическая оптимизация преподавания языков и инженерная лингвистика». Чимкент, 1980. С. 253—254.

Орлов Ю. К. О статистической структуре сообщений, оптимальных для человеческого восприятия // Научно-техническая информация (М.). Сер. 2. 1970. № 2. С. 11—16.

Пиотровский Р. Г., Бектаев К. Б., Пиотровская А. А. Математическая лингвистика. М., 1977.

Сачков Ю. В. Введение в вероятностный мир. М., 1971.

Сиськов В. И. Корреляционный анализ в экономических исследованиях. М., 1975.

Тулдава Ю. А. Проблемы и методы квантитативно-системного анализа лексики. Таллинн, 1987.

Фрумкина Р. М. Роль статистических методов в современных лингвистических исследованиях // Математическая лингвистика. М., 1973. С. 156—183.

Хайтун С. Д. Наукометрия: Состояние и перспективы. М., 1983.

Хастингс Н., Пикок Дж. Справочник по статистическим распределениям / Пер. с англ. М., 1980.

Чупров А. А. Очерки по теории статистики. СПб., 1910.

Шайкевич А. Я. Дистрибутивно-статистический анализ текста: Автореф. докт. дис. Л., 1982.

Шрейдер Ю. А., Шаров А. А. Системы и модели. М, 1982.

Яблонский А. И. Стохастические модели научной деятельности // Системные исследования: Ежегодник. 1975. М., 1975. С. 5—42.

Ястремский Б. С. Некоторые вопросы математической статистики. М., 1960.

Herdan G. Quantitative Linguistics. London, 1964.

Lexis W. Abhandlungen Zur Theorie der Bevölkerungs und Moralstatistik. Vena, 1903.

Lévy P. Calcul de probabilités. Pt. 2, chap. 6. Gauthier; Villars, 1925.

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Как известно, Л. В. Щерба разделял лингвистические исследования на наблюдения и эксперимент. Представленные в предыдущей главе статистические методы применимы в основном к наблюдениям. Однако современная лингвистика становится все более экспериментальной, в особенности в той своей области, которая называется психолингвистикой. Чрезвычайно развиты экспериментальные методы и в фонетических исследованиях. В то же время нет разработок и пособий по обработке лингвистических экспериментов. Исключение составляет изданная малым тиражом книга Т. А. Бровченко и др. (1976), что безусловно отражает факт наибольшей «экспериментальности» фонетики, во всяком случае, в ее щербовском варианте.

Практика, однако, показывает, что применение статистических методов не является тривиальной процедурой. «Столкновение» двух наук требует адаптации их друг к другу. Так, с одной стороны, применение ряда критериев требует «подгонки» материала под определенные рамки. Это естественно, так как всякий статистический критерий моделирует вполне определенную «картину мира». Поэтому, если имеется желание воспользоваться данным критерием, необходимо, чтобы экспериментальный материал в точности соответствовал заданной «картине мира», иначе полученные результаты расчетов отвечают неизвестно чему. Однако требование корректности применения статистических процедур заставляет обычно отбрасывать часть материала, тем самым искажая структуру объекта.

С другой стороны, лингвистический материал вправе требовать адаптации статистического аппарата, разработки специализированных критериев и процедур. Некоторые попытки в этом направлении ведутся, однако они практически отсутствуют, если иметь в виду обработку данных лингвистических экспериментов. Практика показывает, что так называемые стан-

дартные пакеты программ для обработки данных на ЭВМ еще менее пригодны для лингвистического материала. Они учитывают специфику задач и материала (биологического, медицинского или экономического), для которого они построены и который входит в «прикладной» опыт их создателя. Необходимо построение пакета программ для решения собственно лингвистических задач, хотя заранее можно усомниться в его универсальности. Всякий раз задача может быть уникальной. А это значит, что надо адаптировать статистическую процедуру к ней, а не подгонять задачу под описанные каноны.

Учитывая сказанное, остановимся подробнее на тех моментах, где проявляется специфика материала, где применение статистических методов не однозначно.

Статистика изучает варьирующий признак, а признак существует постольку, поскольку существуют его градации. Например, в русском языке предложения могут быть разной длины. Значит, длина — варьирующий признак, а одно, два, три и т. д. слов в предложении — это градации признака «длина». Другой пример — место ударного слога в слове. В русском языке это может быть первый, второй и т. д. слог слова. Это — тоже варьирующий признак, которым может заниматься статистика. Во французском же языке ударение падает всегда на последний слог. Для фонетистов есть смысл говорить о месте французского ударения, хотя бы в сравнении с другими языками, но для статистики это не признак, так как есть лишь одна градация, т. е. нет вариации.

В основе теории вероятностей лежит понятие случайного события (A). Оно сходится с основанием статистики, если иметь в виду под случайным событием появление градаций признака. Но тогда следовало бы говорить о том, что в статистике рассматриваются такие события, для которых справедливо строгое неравенство $0<P(A)<1$, где P — вероятность события A.

Лингвистические признаки могут быть количественными и качественными. Так, количественными являются длина предложения, измеренная в словах, буквах или любых других единицах, глубина фразы (по В. Ингве), число морфем в слове, высота основного тона или интенсивность звука. Ясно, что количественный признак может быть сопоставлен с какой-либо мерой (это непрерывный признак; например, высота основного тона в Гц) или измерен в «штуках» (это дискретный признак; например, число слов в предложении). Отметим, что нас нередко интересуют измерения не только самого признака, но и частоты градаций.

Лингвистические признаки могут иметь и качественную природу — например, «синтаксическая конструкция» (градации могут быть заданы в виде типов синтаксических конструкций; для залога это актив — пассив), «ударность» (можно задать 2 градации — ударный и безударный гласный, а можно задать,

например, 3 градации — ударный, полуударный, безударный), «буква» (их в русском языке 32), «гласная фонема» (их в русском языке 6) и т. д. Ясно, что эти признаки сами не измеряются, можно исследовать лишь одну их характеристику — частоту встречаемости градаций этих признаков. Вопрос этот усложняется тогда, когда исследователя интересует лишь одна градация признака. Например, мы интересуемся предложениями с прямым порядком слов и хотим знать частоту таких предложений в речи. Встает вопрос о том, на каком массиве предложений искать эту частоту — включать ли сюда вопросительные и восклицательные предложения, неполные, сложноподчиненные. По сути, это вопрос о том, какой признак и с какими градациями мы изучаем.

Особые сложности для обработки возникают тогда, когда мы имеем дело с ветвящимся признаком. Например, признак «ритмическая структура слова» — ветвящийся, так как прежде всего это длина слова, а потом — место ударного слога. Можно представить себе, что в условиях некоторых задач все градации такого признака выстраивают в линейку: односложные, хореические, ямбические, дактилические, амфибрахические и т. д. Однако для ряда задач такой строй нелогичен — прежде всего из-за включения сюда односложных слов, строго говоря, не имеющих ритмической структуры. Подобные лингвистические казусы заставляют исследователя, обращающегося к статистическому аппарату, очень четко определять, какой признак и какая градация признака изучается, так как от этого существенно зависят результаты и их интерпретация.

Интуитивно ясно, что если мы хотим по наблюдениям над выборкой сделать какое-то заключение о генеральной совокупности (а именно к этому часто стремятся исследователи), то выборка из нее должна достаточно хорошо (точно) представлять эту генеральную совокупность. Статистики говорят в таком случае о количественной и качественной достоверности выборки. Как это ни странно на первый взгляд, вопрос о количественной достоверности решается проще: если есть какие-то предварительные представления о вариации изучаемого признака, то по соответствующим формулам можно рассчитать необходимый объем, который обеспечил бы заданную точность определения изучаемого параметра, или, другими словами, можно с заданной точностью по выборочным данным сделать какое-то заключение о генеральных параметрах.

Качественная достоверность имеется тогда, когда структура выборки повторяет структуру генеральной совокупности. Например, нас интересует длина слова в газете. Начинающий исследователь поступает зачастую таким образом: берет по 20 строк из передовиц, политических заметок, спортивных новостей и т. д. и на этой объединенной выборке получает значение средней длины слова в газетном тексте. Однако известно,

что в газете рубрики занимают не одинаковый объем. Поэтому полученный описанным способом результат не достоверен; полученный параметр отвечает на какой-то другой, но не на поставленный вопрос. Следовало бы составить выборку так, чтобы был учтен вес (объем) каждой рубрики в газете.

Итак, в случае, когда структура генеральной совокупности известна, точнее — известны факторы, влияющие на изучаемый параметр (в приведенном примере — рубрики), выборку строят в соответствии с этими распределениями. Однако структура генеральной совокупности известна не всегда. В таком случае выборку составляют по случайному закону, т. е. с помощью любого датчика случайных чисел. Например, если надо узнать среднюю длину слова в романе «Война и мир», то можно по таблице случайных чисел выбрать страницы, которые войдут в выборку.

Иногда оба пути оказываются невозможными. Например, нужно изучить характеристики гласных в нормативном произношении. Эти характеристики могут варьировать у разных лиц. Если нет возможности настолько увеличить материал, чтобы в группу дикторов входили люди разного возраста, мужчины и женщины, имеющие разный темп говорения, различные тембральные особенности и т. п., то выбирают принцип представительности. Это значит, что на основании лингвистических соображений (а иногда — чисто интуитивно) выбирают диктора — типичного (или образцового) представителя нормы.

Построение распределения необходимо прежде всего для того, чтобы выявить его форму. Статистические критерии применимы не для всех видов распределений. Прежде всего, распределение должно иметь одну вершину, т. е. быть одномодальным. Если мы получили две или более вершин (при соблюдении закона соответствия объема выборки числу класс-интервалов, или градаций признаков), статистики говорят о том, что наблюдения принадлежат разным генеральным совокупностям. Такую выборку следует внимательно проанализировать и найти способ разбиения ее на два или более множеств, чтобы получить одновершинные распределения. Такие распределения говорят об однородности изучаемого признака.

Итак, генеральная совокупность в статистике — это множество объектов, организованных не просто по изучаемому признаку, а обязательно — по однородному признаку. И здесь возможно противоречие между лингвистическим и статистическим пониманием генеральной совокупности. Разберем следующий пример.

Мы изучаем длительность русских гласных. Безусловно, для фонетиста это признак, однородный на фонемном уровне. Распределение же длительностей гласных оказывается бимодальным — с одной вершиной около 50—60 мс, а с другой — около 110—120 мс. С точки зрения статистики этот признак неодноро-

ден. Нетрудно догадаться, что около первой моды группируются безударные, а около второй — ударные гласные. Если мы хотим дальше применять статистические методы, мы должны подчиниться требованиям математической статистики и рассматривать две генеральные совокупности порознь.

Заметим, что выявление двух или более мод при построении распределения может оказаться и хорошим аппаратом для установления закономерностей, о которых до опыта мы могли ничего не знать, так как не всегда априори можно предположить неоднородность материала в смысле изучаемого признака.

Не так оказывается прост вопрос о средней арифметической, которая знакома всем со школьной скамьи. Сложности возникают прежде всего тогда, когда нужно найти среднюю от средних. Пусть, например, на чтении отрезка текста сосчитаны отдельно средние длительности гласных и согласных. Предположим, что получились такие данные:

гласные	согласные
$n_1 = 40$, $\bar{x}_1 = 90$ мс;	$n_2 = 120$, $\bar{x}_2 = 50$ мс.

Спрашивается: какова средняя длительность звука?

Среднюю можно найти двумя способами. I способ:

$$\frac{\bar{x}_1 + \bar{x}_2}{2} = 70 \text{ мс.}$$

Это так называемая средняя невзвешенная. II способ:

$$\frac{\bar{x}_1 \cdot n_1 + \bar{x}_2 \cdot n_2}{n_1 + n_2} = 60 \text{ мс.}$$

Это — средняя взвешенная. Какая же из них верна?

По-видимому, верны обе средних, но каждая — для своего случая. Средняя невзвешенная нужна для случая, когда мы хотим найти среднюю длительность звука независимо от того, как часто встречаются эти типы звуков в речи. Это — как бы средняя длительность звука в языковой системе. Средняя взвешенная говорит о средней длительности звука в данном звучащем тексте; она зависит от того, как часто встречаются в тексте длинные и короткие звуки. Это — длительность звука в речи. Таким образом, виды средней арифметической коррелируют с противопоставлением речи и языка.

Приведем еще один пример. На пленку записан разговор нескольких людей, т. е. полилог. Пусть в полилоге участвовало 4 диктора. Нужно узнать средний темп речи. Отметим сразу, что заданный таким образом вопрос некорректен. Какой темп речи нас интересует на самом деле? Средний темп дикторов, участвовавших в полилоге? Тогда это средняя невзвешенная. Средний темп звучания данного полилога? Тогда это средняя взвешенная.

В лингвистических работах бытует смешение средней арифметической и медианы, т. е. употребление первой вместо второй. Напомним, что если исследуемый признак лишь кажется количественным (например, в силу того, что применены числовые кодировки типа отметок вместо качественных градаций признака), показателем центральной тенденции должна быть медиана. В принципе, в ряде случаев замена ее на среднюю арифметическую не меняет выводов, однако в этом следует убедиться, произведя несколько параллельных расчетов. Наш опыт показывает, что при расчете факторного анализа по экспертным оценкам в одних случаях результаты для средних и медиан не расходятся, в других же выводы оказываются разными.

При обработке экспериментальных данных зачастую встает вопрос о существенности тех или иных ответов, что обычно сводится к оценке отклонения эмпирической частоты (f) от теоретической вероятности (p). Например, аудиторам предлагается прослушать вырезанный из магнитной записи гласный звук и опознать его. Испытуемым сообщается, что возможны 6 вариантов ответа (по числу гласных фонем). Пусть в результате эксперимента получаем, например, такие частоты реакций:

V	а	о	u	е	ы	i
$f(V)$	0,30	0,19	0,25	0,10	0,16	0,00

Обращает на себя внимание тот факт, что ответы |а| и |u| имеют максимальные частоты. Обычно неслучайность каждого ответа определяется по t-критерию Стьюдента. Однако для этого требуется задать теоретическую вероятность (исходную, априорную), т. е. оценить доопытное знание испытуемыми возможностей появления различных стимулов. Это можно сделать несколькими способами.

1) Поскольку все ответы вводятся непосредственно перед экспериментом, полагаем, что все они равновозможны. И тогда $p|а|=p|u|=\ldots=0{,}17$. В условиях данной задачи интересующие нас ответы |а| и |u| оказываются случайными.

2) Хотя было введено 6 ответов, реакция |i| практически невозможна, так как гласный в действительности стоял после твердого согласного. Это подтверждается и $f|i|=0$. Можно думать, что испытуемые после предъявления им нескольких гласных стимулов уточняют этот алфавит и работают в поле из 5 ответов. Тогда априорная $p=0{,}20$, и неслучайным оказывается лишь ответ |а|.

3) Несмотря на введение алфавита ответов они в действительности не являются равновозможными, и испытуемые пользуются теми частотами, которые приобретены ими в речевом опыте. Тогда $p|а|=0{,}42$, $p|u|=0{,}08$ (Штерн, 1988, с. 206), и неслучайным является ответ |u|.

Несколько усложним теперь задачу. Пусть нас интересует в действительности огубленность гласного в ответах. Тогда часто-

та огубленных равна 0,44. Если исчислять тремя предложенными способами вероятности |о| и |u|, то получим, что в первом и третьем случаях ответы огубленных неслучайны, а во втором — в пределах случайного. Однако можно предположить еще один способ расчета: можно предположить, что испытуемые вначале реагируют на огубленность гласного, а лишь затем выбирают одним из способов конкретную реализацию. Тогда на первой ступени p(ог.) $=0{,}5$ (либо другая p). И тогда ответы в смысле огубленности оказываются случайными.

Тогда следует сделать вывод о том, что результаты расчетов (и соответственно выводы относительно свойств стимулов) будут зависеть от того, какую процедуру примем мы в качестве исходной, т. е., строго говоря, от некоторого «экстрастатистического момента».

Исследуя только «пики» распределений, «моды», мы теряем большую часть информации, поскольку при этом виде анализа остаются «за бортом» все ответы, которые с точки зрения статистического критерия оказываются случайными. В то же время два стимула могут иметь одинаковые «пики» в ответах и различаться распределениями в целом («фонемными изображениями стимулов», по Л. А. Чистович). Поэтому нужно применять критерии сравнения распределений.

Однако ситуация осложняется в тех случаях, когда алфавит ответов не ограничен. В этом случае, впрочем, как и в предыдущем, может быть применена с успехом оценка энтропии опыта (H) по Шеннону. Она обладает тем преимуществом, что позволяет уйти от вопросов об априорном распределении частот реакций и об алфавите ответов. Правда, при этом мы уходим от проблемы «качества» реакций. В результате оценивается неопределенность опыта, или неопределенность стимула в перцептивном эксперименте.

Следует предостеречь исследователей от возможных ошибок при вычислении усредненных оценок энтропии. Важно различать два способа. В одном случае можно получить энтропию на каждый стимул, а затем усреднить ее для всей группы стимулов — $\bar{H}$. В другом случае можно получить среднее фонемное изображение группы стимулов, а затем для него рассчитать энтропию — $H\{\bar{p}\}$. Эти оценки не совпадают: $\bar{H}<H\{\bar{p}\}$. Степень расхождения между обеими H тем больше, чем больше варьируют фонемные отображения внутри группы стимулов; равенство возможно лишь в том случае, когда фонемные изображения всех стимулов группы совпадут. Покажем это на примерах.

Пример 1

	а	о	u	е	і	ы
S_1	50	50				
S_2	50	50				
S_3	50	50				

Пример 2

	а	о	u	е	і	ы
S_1	50	50				
S_2			50	50		
S_3					50	50

В примере 1 неопределенность каждого стимула S равна 1 биту и $\overline{H}=1$. Среднее распределение ответов (R) — {50, 50, 0, 0, 0, 0} и $H\{\overline{p}\}=1$. В примере 2 энтропия каждого стимула также равна 1 и $H=1$. Но усредненное распределение стимулов — {17, 17, 17, 17, 17, 17}. И тогда $H\{\overline{p}\}=2{,}58$ бита. Следует учитывать качественное истолкование каждой из них. $\overline{H}$ говорит о средней неопределенности группы стимулов, а $H\{\overline{p}\}$ — о неопределенности усредненного стимула из некоторой группы с учетом вариации фонемных изображений.

Заметим также, что H, если имеется много разных стимулов, может рассматриваться как количественный признак и, таким образом, исследоваться с помощью распределений и т. п. методов для количественных признаков. Решение некоторых задач перцептивной фонетики с помощью такого способа см. в работе Л. В. Игнаткиной, А. С. Штерн (1984), а исследование ассоциативной силы русского слова представлено в работе И. Г. Овчинниковой, А. С. Штерн (1989).

Исследования показали, что весьма плодотворным является применение дисперсионного анализа к результатам как экспериментов в области фонетики, так и экспериментов на ассоциации, толкование и др. Следует, однако, заметить, что имеются известные трудности в применении этой процедуры (Штерн, 1992). Прежде всего, применение дисперсионного анализа является корректным, если соблюдено требование ортогональности (независимости) факторов. Однако язык — принципиально неортогональная система. Так, например, в русском языке неортогональны признаки шумность — сонорность и звонкость — глухость, так как все сонорные — фонологически звонкие, и т. п. При подборе словесного материала оказывается, что длинных высокочастотных слов в английском языке мало (или они вовсе отсутствуют, что, впрочем, известно лингвистам) или, например, отсутствуют односложные глаголы в немецком языке и т. п. Это, с одной стороны, как бы искажает объективную картину действия факторов, но, с другой стороны, отражает специфику языка как системы.

Распространение дисперсионного анализа в лингвистических исследованиях неожиданно натолкнулось на психологическое «неприятие» результатов со стороны некоторых исследователей, поскольку веса факторов (η_x^2) оказываются невелики (например, для восприятия слова они не превышают в основном 10—15%, часто оказываются равными лишь 1—2%, что, впрочем, по F-критерию Фишера оказывается значимым, а сумма весов факторов, как правило, не превышает 35%). Однако небольшое значение корреляционного отношения при условии его существенности может соответствовать существенности разницы по t-критерию Стьюдента на 5%-ном уровне при довольно неболь-

шой разнице между средними (и соответственно малом η_x^2). Однако тот факт, что критерий Стьюдента давно (хотя, к сожалению, в ряде случаев некорректно) применяется в лингвистике, не смущает исследователей, соответствующее же ему небольшое корреляционное отношение непривычно. Из накопленного опыта следует лишь, что язык — весьма многофакторная и плохо формализуемая система.

Коэффициент корреляции может стать орудием для нахождения единиц принятия решения, в которых работает человек при восприятии речи. (То же можно сказать и о дисперсионном критерии, где в случае линейной зависимости корреляционное отношение равно квадрату коэффициента корреляции.) Для некоторых случаев известно, что должна быть корреляция между признаком и результатами восприятия. Так, широко подтверждено на экспериментах разной модальности, что при восприятии происходит вероятностное прогнозирование, т. е. чем частотнее стимул, тем лучше (легче, правильнее и т. п.) он воспринимается. Однако сам признак «частотность» — исходно непрерывный (практически непрерывна частотность слов по частотному словарю). Можно задать несколько способов разбиения диапазона частот на градации. Например, в линейной шкале, в логарифмической, при более сложных (смешанных) способах. Разным может быть и число градаций. Для каждого способа можно получить свой коэффициент корреляции с частотой правильного опознания (или свое корреляционное отношение). Максимальный коэффициент в таком случае будет соответствовать оптимальному заданию класс-интервалов; оптимальному — в смысле единиц этого признака.

При работе с коэффициентами корреляций наибольшее число вопросов возникает по трактовке этих коэффициентов: что является большой, а что малой корреляцией и т. п. Существует, скажем, шкала Гилфорда, где предлагается очень высокой считать корреляцию от 0,8 до 1,0, а средней — от 0,4 до 0,6. Но эти трактовки предложены для психологических и педагогических исследований. Не следует забывать о специфике лингвистических признаков и языка как системы. Может оказаться, что при очень большом числе экспериментов однотипные корреляции не превышают 0,7. По-видимому, в таком случае, исходя из природы признака, следует изменить шкалу и считать, что 0,6—0,7 — это очень высокая корреляция, а 0,4—0,6 — высокая и т. д.

* *
*

В заключение хочется отметить, что не сам по себе статистический аппарат и даже не взаимная адаптация методов, материала и аппарата друг к другу представляет наибольшие трудности. Оказывается, что всего сложнее для исследователя

увидеть в своей лингвистической гипотезе статистическую задачу, правильно подобрать статистическую процедуру, а затем дать адекватное толкование статистическим расчетам в терминах лингвистической проблемы. К сожалению, подобных методических разработок пока нет (как, впрочем, нет их и в других прикладных областях). Пока единственный для исследователя путь — равное владение как лингвистической проблематикой, так и аппаратом математической статистики.

Литература

Бровченко Т. А., Варбанец П. Д., Таранец В. Г. Метод статистического анализа в фонетических исследованиях. Одесса, 1976. 100 с.

Игнаткина Л. В., Штерн А. С. Неопределенность при восприятии гласных // Слух и речь в норме и патологии. Вып. 5. Л., 1984. С. 3—10.

Носенко И. А. Начала статистики для лингвистов. М., 1981. 157 с.

Овчинникова И. Г., Штерн А. С. Ассоциативная сила русского слова // Психолингвистические проблемы фонетики и лексики / Ред. А. А. Залевская. Калинин, 1989. С. 110—117.

Штерн А. С. 1) Некоторые статистические характеристики русской спонтанной речи // Фонетика спонтанной речи / Ред. Н. Д. Светозарова. Л., 1988. С. 196—224; 2) Перцептивный аспект речевой деятельности (экспериментальное исследование). СПб., 1992. 236 с.

СЕТЕВОЕ ПРЕДСТАВЛЕНИЕ ЛИНГВИСТИЧЕСКИХ ДАННЫХ

Информация, используемая в лингвистическом процессоре, структурирована определенным способом, в большой степени определяющим содержание процедур организации лингвистических данных и поиск нужных элементов. Будем называть его способом представления данных, или просто представлением данных. В представлении лингвистических данных различаются субстанциональный компонент и реляционный компонент. Разделение этих типов проводится на основе понятия лингвистическая конструкция, под которой понимается составной (комплексный) фрагмент лингвистического описания. Элементы этого фрагмента рассматриваются как неделимые объекты, связанные определенными отношениями. Элемент конструкции, в свою очередь, может представляться в виде конструкции, являясь встроенной структурой. В лингвистической конструкции могут представляться такие отрезки текста, как, например, структура морфов в слове. В этой конструкции элементами являются алломорфы, входящие в слово, а отношением является их линейный порядок в слове. Между алломорфами морфа могут быть установлены отношения, например, частоты встречаемости в текстах, тогда они также формируют лингвистическую конструкцию парадигматического типа. Та информация, которая представлена в виде элементов лингвистических конструкций, составляет субстанциональный компонент, соответственно отношения между элементами конструкций формируют реляционный компонент лингвистических данных.

Способ представления лингвистических данных в автоматизированной системе отражает противоположение субстанционального и реляционного компонентов лингвистической информации в некоторой абстрактной структуре, которая является определяющей для способа представления и название которой

чаще всего используется для наименования типа представления. Сетевое представление данных, следовательно, — задание лингвистической информации в виде сети. Абстрактная структура типа «сеть» — это ориентированный псевдограф, в котором непустое множество вершин V и множество E упорядоченных пар вершин $(a, b) \in E$, называемых дугами графа, отображаются в множество меток вершин P и множество меток дуг T. Среди множества дуг E есть дуги вида (a, a), которые называются петлями, и может быть более одной дуги вида (a, b), это так называемые параллельные дуги. Множества меток P и T совместно с множеством дуг E позволяют представлять субстанциональный и реляционный компоненты лингвистической информации процессора.

Широкое применение сетевых представлений при обработке информации обусловлено несколькими причинами. Среди преимуществ сетевых представлений наиболее очевидны их способность отражать структурные отношения сложных лингвистических структур и операциональное удобство поиска информации. Первое преимущество обусловлено отсутствием ограничений структуры массива информации, так как более специальные виды структур могут вступать в противоречие с реально встретившимся в лингвистической системе отношением. Второе преимущество состоит в упрощении процедуры выбора гипотез, часто выполняемой в лингвистических задачах. Сведение вместе всех отношений некоторого элемента лингвистической структуры позволяет оперативно оценить набор альтернатив анализа в заданной точке текста и активно выдвигать гипотезы о структуре последующих фрагментов текста. Важной особенностью сетевых представлений является то, что субстанциональный компонент данных может быть размещен как в виде меток вершин, так и в виде меток дуг. В зависимости от этого мы будем различать стационарные сети и динамические сети (сети переходов).

Наряду с перечисленными выше позитивными характеристиками сетевых представлений лингвистической информации имеется и ряд негативных черт. Это, например, громоздкий способ отображения лингвистических конструкций; отсутствие четких границ отображаемых конструкций в сетях; сложность разграничения типов конструкций в сетях, объединяющих несколько лингвистических конструкций. Таким образом, наиболее существенным моментом при работе с сетями является определение способа фиксации границ лингвистических конструкций в представлении. Отображение лингвистической конструкции в сетях будем называть сетевой конфигурацией, или просто конфигурацией.

Определяющим элементом лингвистических сетевых представлений являются конфигурации, отображающие лингвистические конструкции, необходимые для анализа высказываний

естественного языка. Одно и то же лингвистическое описание может быть представлено с разной степенью детализации несколькими лингвистическими конструкциями, а следовательно, и конфигурациями в сети. Рассмотрим в качестве примера несколько синонимических рядов:

1) безопасный, неопасный, безобидный;
2) невинный, безобидный, безвредный;
3) невинный, невиновный, неповинный, невиноватый, правый, безвинный;
4) справедливый, правый, праведный, законный;
5) справедливый, верный, достоверный;
6) меткий, точный, верный;
7) надежный, верный.

На рис. 1 изображен фрагмент сети синонимических рядов. Лингвистическими конструкциями для представления в этой

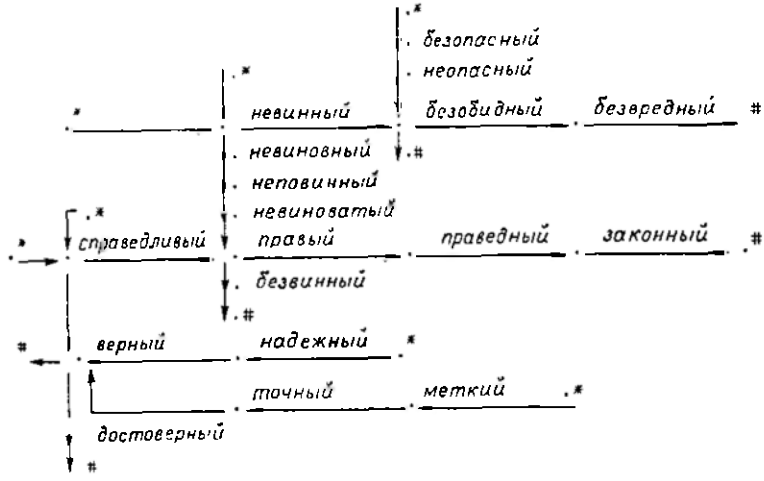


Рис. 1.

сети являются приведенные синонимические ряды. Элементами конструкции являются лексемы. Объединение лексем в ряд производится по некоторому значению — семантической доминанте ряда. Упорядочение лексем в ряду производится по нескольким признакам, в число которых входят точность выражения семантической доминанты, стилистическая окраска, употребительность лексем и т. п. Лексемы синонимических рядов представлены в сети в виде меток вершин. Порядок следования лексем в ряду отражается последовательностью дуг (называемой «путь») от специальной вершины с меткой * (начало пути — начало ряда) до вершины с меткой # (конец пути — конец ряда). Использование этих специальных вершин необходимо, чтобы правильно находить границы ряда.

Выделяемые в синонимической сети пути от начала к концу могут отличаться от первоначально заданных конструкций, например: *невинный, невиновный, неповинный, правый, праведный, законный.* Такие конфигурации будем называть ложными. Для защиты сетевого представления от выделения в нем ложных конфигураций можно использовать два способа. Первый способ состоит в том, что все дуги конфигурации помечаются специальным образом, например, номером конфигурации (рис. 2).

Рис. 2.

Второй способ состоит в преобразовании конфигураций сети таким образом, чтобы конец конфигурации связывался с ее началом специальной граничной дугой, при этом конфигурация типа «путь» преобразуется в конфигурацию «контур» (рис. 3). Дуга, связывающая конец и начало ряда, выделена специальным символом «0».

Первый способ уточнения пределов конфигураций сети называется разметкой, второй — замыканием. Разметка сети является более универсальной операцией, так как применима к конфигурации любого вида, однако она занимает большую часть информационного пространства — информации, хранящейся в сети в виде меток.

В конструкциях приводимого фрагмента сети члены синонимических рядов упорядочены в соответствии с несколькими признаками. Пометы стилистической окраски лексем и их употребительность частично указаны при перечислении элементов синонимического ряда в словаре. Отразить эту информацию в сети можно, дополнив состав меток соответствующими призна-

ками типа *разг., книжн., устар.* и т. п. Эта информация может относиться как ко всем вхождениям лексемы в синонимические

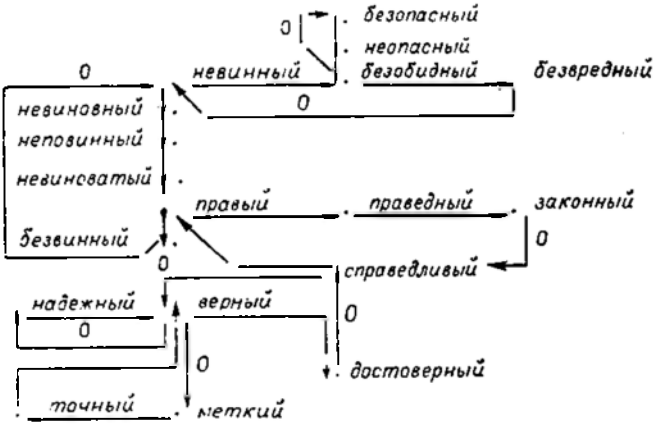


Рис. 3.

ряды, так и к отдельным вхождениям, поэтому желательно заносить ее в сеть в виде меток дуг. Часть этой информации можно представить в структуре конфигураций следующим образом. Ряд синонимов разбивается на три группы: нейтральная

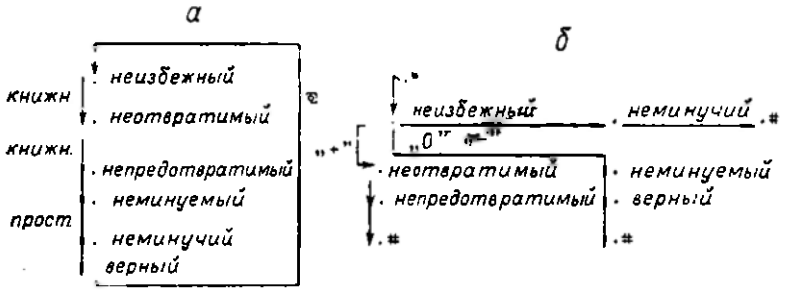


Рис. 4.

лексика, стилистически сниженная лексика (просторечные, разговорные, диалектные выражения), «элитарная» лексика (официальные, книжные выражения). Каждая из групп ряда отображается в конфигурации путем от вершины — основного представителя синонимического ряда. Тогда в зависимости от однородности или неоднородности стилистического состава лексики ряда он будет отображаться в сети либо путем, либо деревом.

На рис. 4 изображены два варианта конфигураций (*а* — «контур», *б* — «дерево») для синонимического ряда *неизбежный, неотвратимый, непредотвратимый, неминуемый, неминучий, верный*. Метка «—» на дуге сети (рис. 4, *б*) указывает на подгруппу синонимического ряда с пониженным статусом лексем, метка «+» указывает на подгруппу повышенного стилистического статуса, метка «0» — на группу нейтральной лексики.

Из приведенного примера видно, что некоторая часть реляционного компонента лингвистической информации может быть представлена в огрубленном виде структурой конфигураций. Использование структурной базы сети для передачи смысловых отношений затрудняет применение структурных методов ограничения конфигураций, например замыкания. Множество дуг сети *E* называется структурой сети, или структурной базой сети. Метки дуг *T*, передавая лингвистическую информацию, выполняют две основные функции: делимитативную и репрезентативную. Основной является делимитативная — функция ограничения (выделения) конфигураций в сети. Вторая функция структуры сети реализуется тогда, когда часть реляционного компонента представлена при помощи дуг.

Лексемы в синонимическом ряду объединены как эквивалентные (или почти эквивалентные) по отношению к основной компоненте их значения, несмотря на различия в оттенках значений, стилистической окраске, сфере употребления и т. п. Слово, с которого начинается ряд, выступает в роли доминанты. Такие ряды являются парадигматическими. Сети, в которых отображаются лингвистические конструкции такого типа, называются парадигматическими.

На рис. 5 приведены два варианта конфигураций (*а* — «дерево», *б* — «контур»), отображающих парадигматическую конструкцию для словоизменения лексемы *путь*. На дугах помещены обозначения морфологических категорий, ассоциируемых со словоформой: *Е* — единственное число, *М* — множественное число, *И* — именительный падеж, *Р* — родительный падеж, *Д* — дательный падеж, *В* — винительный падеж, *Т* — творительный падеж, *П* — предложный падеж.

В данных сетях помимо собственно языковых меток (словоформ) используются их метаязыковые описания, которые называются классификаторами. При первичном заполнении сетей классифицирующая информация должна быть подготовлена и введена в сеть, полученные сети являются классифицирующими. Вместо классификаторов в сетях могут использоваться расставляемые программным путем метки, которые называются маркерами. В сетях с маркерами производится внутренняя классификация связей между фрагментами текста, т. е. вводятся элементы самообучения. В тех случаях, когда и сами связи, дуги сетевого представления, устанавливаются программно, сети называют обучающимися.

Первая конфигурация (см. рис. 5, *а*) позволяет для лексемы *путь* и классификационной метки (например, *МД*) найти представление словоформы (*путям*). Во второй конфигурации для словоформы (*пути*) устанавливается исходный вид лексемы

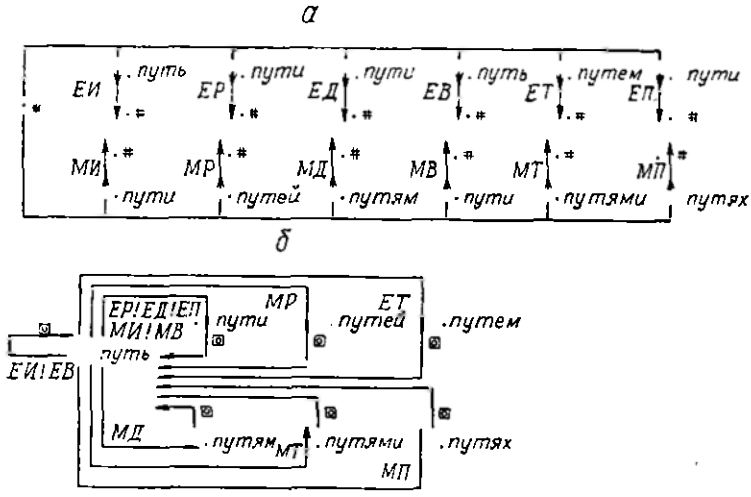


Рис. 5.

(*путь*) и определяются обозначения одной или нескольких морфологических интерпретаций словоформы (*ЕР*, *ЕД*, *ЕП*, *МИ*, *МВ*). Информация, которая используется в качестве условия поиска, определяет структуру конфигураций: количество вершин, направление дуг. Условие поиска может задаваться как метками вершин, так и метками дуг; совокупность меток, используемых для анализа, будем называть поисковыми признаками. Поисковые признаки приобретают определенный смысл после определения стратегии анализа текста. При восходящем анализе «снизу вверх» (см. рис. 5, *б*) из текста выделяются фрагменты, которые составляют часть поисковых признаков; они дополняются классификационными метками, которые были получены на предыдущих стадиях анализа. Нисходящая модель анализа, или анализ «сверху вниз» (см. рис. 5, *а*), предполагает наличие генератора гипотез, заранее настроенного на определенный объем текстов. Генератор, пользуясь имеющимися у него приоритетами классификационных меток, выдвигает гипотезу о том, какой фрагмент находится в данной точке текста. Гипотеза проверяется, и в случае несовпадения предполагаемого и реального фрагментов выдвигается новая гипотеза.

Поисковые признаки проецируются на структуру сети, в ко-

торой их корреляты, вершины и дуги, составляют актуализированный компонент сети. Если отображение V на P и E на T взаимно-однозначно (т. е. все метки различаются), то в такой простой сети количество элементов актуализированного компонента равно числу поисковых признаков (см. рис. 5, *б*). В противном случае сеть будет сложной, в ней число актуализированных элементов сети может превосходить число поисковых признаков (см. рис. 5, *а*). Для каждого из элементов актуализированного компонента определяется зона поиска, которая может задаваться минимальным или полным способом в зависимости от принятого в данном лингвистическом процессоре описания связей вершин в сети. В минимальном варианте зоной поиска для вершины являются выходящие из нее дуги, для дуги — вершина, в которую она входит. В полном варианте зоной поиска для вершины являются все дуги, которые входят и выходят из вершины, а для дуги — обе вершины, соединяемые ею. Актуализированный элемент сети вместе с зоной поиска образует подсеть, которую будем называть элементарным пространством поиска. В зависимости от способа соединения поисковых признаков в условии поиска из элементарных пространств формируется пространство поиска. Пространство поиска может расширяться, для того чтобы восстановить пропущенные в тексте фрагменты. В сформированном поисковом пространстве выбор нужных элементов — решение задачи лингвистического анализа — может осуществляться разными способами в зависимости от правила принятия решения, или решающего правила. Простейшее решающее правило состоит в выборе меток вершин и дуг из поискового пространства, отличных от поисковых признаков. Это правило простого дополнения. Например, в приведенной выше конфигурации (см. рис. 5, *в*) если известна словоформа *путем*, то можно установить ее исходную форму и морфологическую метку. Более сложные, решающие правила могут потребовать выделения конфигураций определенного вида в пространстве (конфигурационное правило), перестройки поискового пространства или комбинирования конфигураций (фреймовые правила).

Рассматриваемые выше примеры относились к парадигматическим сетям. Наряду с ними широко используются синтагматические сети, в конструкциях которых объединяются одновременно встречающиеся в тексте фрагменты. Объединение в конструкцию элементов проводится по нескольким принципам. Во-первых, элементы конструкции могут располагаться непосредственно друг за другом в тексте или речи, т. е. реляционным компонентом конструкции является отношение линейного следования элементов; такие сети будем называть линейными. Во-вторых, в конструкции может отражаться сочетаемость в виде модели управления или лексической сочетаемости. Они фиксируются в сетях сочетаемости.

Приводимая выше парадигматическая конфигурация для словоизменения слова *путь* в синтагматической сети будет описываться как набор из семи конфигураций (рис. 6), в которых представлено сочетание «основы» слова и «окончания».

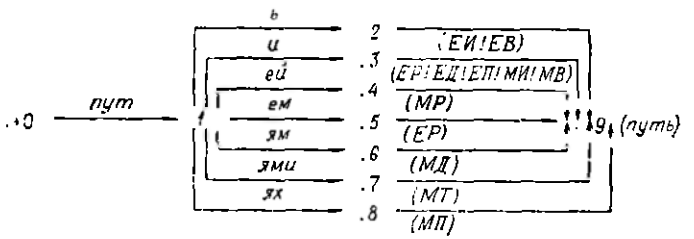


Рис. 6.

На рис. 6 от вершины «и» ведет дуга, помеченная множеством обозначений морфологических форм, реально это дизъюнкция меток, заменяющая набор параллельных (однонаправленных) дуг. Для выделения в их составе групп меток введем понятие пучка дуг. Пучок дуг — это группа параллельных дуг, имеющих один и тот же компонент в метке. Этот компонент является доминантой пучка. На рис. 6 от вершины «и» можно провести два пучка, доминантой первого будет метка *Е* (единственное число), доминантой второго — *М* (множественное число). Дуги в пучке должны быть различимы, т. е. должны быть различающие их метки, которые служат дифференциаторами.

Насыщенность связей в сети можно оценить, усреднив локальные степени вершин — полустепени исхода и захода вершин (количество входящих в вершину и выходящих из нее дуг). Число выходящих дуг будем называть локальным ветвлением. Среднее значение ветвления по всем вершинам будем называть средним коэффициентом ветвления сети. Если значение этого коэффициента не превосходит 2, то сеть является редкой. Алгоритмы обхода редких сетей наиболее эффективны. Содержательно этот факт можно проинтерпретировать следующим образом: редкая сеть имеет высокую предсказывающую силу. При организации лингвистических сетей необходимо ориентироваться на структуры такого типа.

Первым способом понижения ветвления в сети является задание связей сильно ветвящихся вершин при помощи специальных меток — моделей. Второй способ обеспечивается введением в сеть специальных вершин-заместителей, имеющих классификационные метки. В качестве примера рассмотрим представление словоизменения лексемы *путь*, используя введенные А. А. Зализняком типы субстантивного склонения *См8в*, где *С* указывает

на субстантивное склонение, *м* — мужской род, *8* — подтип склонения, *в* — перенос ударения на флексию (рис. 7).

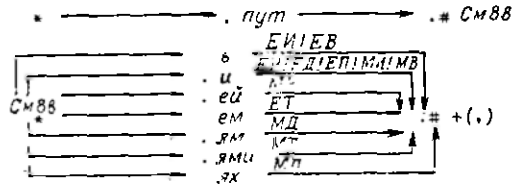


Рис. 7.

В приведенном на рис. 7 примере особая метка + (,) обозначает процедуру, состоящую в конкатенации меток (добавлении к основе *пут* соответствующего окончания). Такие метки называются процедуральными. Они могут представлять собой достаточно сложные программы.

Важное противоположение типов сетей связано с характером отображения объектов сети в множество элементов субстанционального компонента лингвистических конструкций. В предыдущих примерах эти элементы выступали в качестве меток вершин (стационарные сети). В сетях переходов они выступают в качестве меток дуг. На рис. 8 показана сеть переходов для представления словоизменения словоформы *путь*.

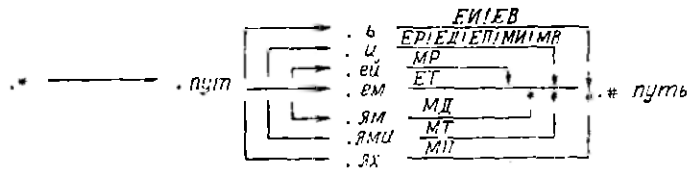


Рис. 8.

В качестве меток вершин выступают так называемые номера состояний, среди которых выделяются две группы: начальные состояния (они соответствуют входным вершинам стационарных сетей) и конечные состояния (они сходны с выходными вершинами). Начальные состояния отмечены символами +, конечные состояния — символами —.

Переход от вершины к вершине в сети переходов происходит при условии совпадения фрагмента анализируемого текста с меткой дуги, т. е. пространство поиска всегда является локальным. Обнаружение конца конфигурации фиксируется по достижению конечного состояния. В круглых скобках под ду-

гами указываются процедуральные метки, которые в данном случае снабжают необходимой для анализа информацией: указывают морфологическую характеристику словоформы и исходную форму лексемы.

Усложненный вариант сетей переходов — рекурсивные сети имеют дополнительное свойство: они позволяют переносить анализ в другую часть сети и при достижении конца конфигурации в ней снова возвращаться на место вызова.

В сетях переходов эффективно фиксируются именно линейные последовательности элементов конструкций. Различные виды сетевых представлений позволяют настраиваться на нужные типы лингвистической информации: парадигматические ряды, линейные и структурные связи фрагментов текста. Использование в сетях классификационных и процедуральных меток дает возможность сочетать элементы декларативного и процедурального представления. Подбор типа сетевого представления для решения задач обработки текста в лингвистическом процессоре должен сочетаться с выбором стратегии анализа текста, которая обусловлена качеством входной текстовой информации и требованиями, предъявляемыми к результатам преобразования информации в лингвистическом процессоре.

Литература

Азарова И. В. Использование сетевых представлений лингвистических данных при автоматической обработке текста: Дис. ... канд. филол. наук. Л., 1982.

Вудс В. А. Сетевая грамматика для анализа естественного языка // Кибернетический сборник. Нов. сер. Вып. 13. М., 1976. С. 120—158.

Минский М. Фреймы для представления знаний. М., 1979.

Харари Ф. Теория графов. М., 1973.

Цейтин Г. С. Программирование на ассоциативных сетях // ЭВМ в проектировании и производстве. Л., 1985. С. 116—148.

Bobrow D. G., Winograd T. An Overview of KRL, a Knowledge Representation Language // Cognitive Science. 1977. Vol. 1, N 1. P. 3—47.

Schubert L. K. Extending the Expressive Power of Semantic Network // Artificial Intelligence. 1976. Vol. 7, N 2. P. 163—198.

ПСИХОЛИНГВИСТИКА

1. **Из истории становления психолингвистики.** Начиная с работ Ф. де Соссюра в лингвистике стали разрабатываться идеи, связанные с пониманием языка как структуры, языка, существующего вне человека. Однако, достигнув замечательных успехов в выявлении принципиальных закономерностей устройства языка как системы знаков, структурная лингвистика встала перед невозможностью решить ряд проблем, связанных с тонкостями реального функционирования языка в процессе общения людей. Именно это заставило ученых вновь обратиться к человеку говорящему. Исследователи поставили задачу изучения процессов говорения и процессов понимания, или восприятия, речи. Так в середине XX в. возникла новая наука — психолингвистика. Основной объект ее исследования — это механизмы речевой деятельности, т. е. процессы говорения и восприятия. В связи с этим встают специфические вопросы: каковы единицы порождения и восприятия речи? каково становление этих механизмов в онтогенезе? как соотносятся лингвистические и психолингвистические единицы? какова психологическая реальность лингвистических моделей? — и т. д., и т. п.

Следует отметить, что интерес к психологическим аспектам речевой деятельности людей возник в новое время отнюдь не в 50-х годах XX в. Психологическое направление, основывающееся как на успехах психологии в XIX в., так и на развитии сравнительно-исторического языкознания, дало новый, удивительный взгляд на язык, на исторические изменения языка, на роль человека в них. Многие идеи, активно эксплуатируемые современной теоретической психолингвистикой, прежде всего — тезис о языке как деятельности, были заложены уже в трудах В. фон Гумбольдта, а языкознание второй половины XIX в. было, по существу, психологическим языкознанием (см. Звегинцев, 1964). Из западноевропейских традиций изучения речемыс-

лительной деятельности в первой половине XX в. необходимо отметить французскую школу Ж. Пиаже. Однако психолингвистика второй половины XX в. — это принципиально новый этап развития науки. В этой сравнительно молодой дисциплине можно выделить несколько направлений, определяемых как лингвистическими, так и, в особенности, психологическими позициями исследователей.

Психолингвистика как новое научное направление официально заявила о себе в 1953 г., когда в Блумингтоне (США) состоялся знаменитый междисциплинарный семинар, итоги работы которого были опубликованы в 1954 г. в книге «Psycholinguistics. A survey of theory and research problems», вышедшей под редакцией известных американских ученых — психолога Ч. Осгуда и антрополога и этнографа Т. Сибеока. Направление психолингвистики, которое оформилось благодаря этой книге, можно условно назвать «ассоцианистским». Среди его истоков были не только идеи психологического направления бихевиоризма (поведенческой психологии), но и идеи американского структурализма — дескриптивной лингвистики (которая, в свою очередь, начиналась с осмысления бихевиористской модели «стимул — реакция»).

Возникший у ученых активный интерес к психолингвистике уже к середине 60-х годов был направлен на новое, «трансформационистское», психолингвистическое направление, связанное прежде всего с именами известных американских ученых — психолога Дж. Миллера и лингвиста Н. Хомского. Здесь проявилось влияние генеральной линии развития лингвистической мысли второй половины XX в. — развитие интереса к синтаксису предложения. Это направление базировалось на генеративной грамматике Н. Хомского, получившей в то время чрезвычайную популярность среди лингвистов. Более четкая лингвистическая ориентация этого психолингвистического направления и интерес к моделям Н. Хомского, а также, разумеется, то, что с исследования связей слов и/или групп слов центр тяжести перемещается на предложение, — все это заметно оттеснило ассоцианистское направление на периферию психолингвистических исследований. А доминирующее положение школы Дж. Миллера — Н. Хомского привело к тому, что именно с этой школой стало отождествляться понятие психолингвистики как таковой.

Наконец, новый виток развития лингвистической мысли в 70-х годах, приведший к бурному развитию лингвистики текста, заставил переориентироваться и значительную часть психолингвистики. Развитие же когнитивной психологии, ориентированной на моделирование структуры знаний, процессов приема, переработки, хранения и выдачи информации человеком (во многом в связи с компьютерным моделированием искусственного интеллекта, т. е. с так называемой компьютерной метафорой), насытили научную среду новыми принципиальными и гло-

бальными идеями, что и привело к повороту, к созданию глобальных теорий и попыткам построения более или менее детально проработанных моделей структурирования текста.

Такова внешняя канва развития современной психолингвистики. Однако необходимо сделать существенное дополнение.

Отечественная психолингвистика опирается не только на достижения зарубежных направлений, но и на богатые традиции советской науки. В психологическом аспекте она базируется прежде всего на культурно-исторической концепции в психологии 20—30-х годов Л. С. Выготского, включающей в себя два основных момента: 1) речевая деятельность есть именно деятельность, и, следовательно, она включает в себя мотив, цель, иерархическую структуру и т. п.; 2) человек есть существо социальное, и, следовательно, социум формирует и регулирует его деятельность. Не случайно поэтому многие советские исследователи термину «психолингвистика» предпочитают термин «теория речевой деятельности».

В лингвистическом же аспекте отечественная психолингвистика опирается прежде всего на традиции Л. В. Щербы. По существу, все основные концептуальные положения советской психолингвистики четко сформулированы в его основополагающей статье «О трояком аспекте языковых явлений и об эксперименте в языкознании»: о важности изучения процессов говорения и процессов понимания как одного из аспектов языковых явлений, об опоре на «отрицательный» языковой материал, в особенности — на детскую речь и патологию речи, о важности развития эксперимента в языкознании. Среди предшественников советской психолингвистики, конечно, нужно указать и М. М. Бахтина (Волошинова), и Л. П. Якубинского, и Е. Д. Поливанова. Наконец, за знаменитой полемикой ленинградской и московской фонологических школ, в конечном счете, скрывается борьба принципов лингвистики, опирающейся на психологические механизмы речемыслительной деятельности (ленинградская школа), и принципов лингвистики, опирающейся на структуралистски-антипсихологические установки (московская школа).

Поворот в последние десятилетия к человеку, учет «человеческого фактора», характерный для всех гуманитарных наук, коснулся и лингвистики. И здесь среди многих направлений современной лингвистики психолингвистика продвинулась дальше других в своих исходных посылках, ибо в явном виде в качестве основополагающего она берет тезис об опоре на психологию человека, об учете этого аспекта в исследовательской практике и в теоретическом моделировании.

В этом плане бесспорно сильным моментом для психолингвистики является резкое расширение фактографической базы исследования. Это расширение связано, во-первых, с углубленным изучением не только «нормальных» проявлений в речевой

деятельности, но и «отрицательного» материала и, во-вторых, с широким использованием наряду с традиционным методом — методом-наблюдением — разнообразных методик экспериментального исследования.

На особенностях методов и материала психолингвистических исследований следует остановиться более подробно.

2. **Методы психолингвистических исследований.** В психолингвистических исследованиях применяются в качестве методов и наблюдение, и эксперимент. (Как будет показано ниже, и наблюдение, и эксперимент используются как в нормальной коммуникации, так и при различных отклонениях от нее.) При проведении наблюдения исследователь только фиксирует естественное общение людей (на магнитофон, на видеопленку, на бумагу) либо пользуется естественной авторской фиксацией (сочинениями, дневниками, письмами и т. д.). При эксперименте исследователь сам активно формирует некую коммуникативную ситуацию, чаще всего включающую в себя эксплицитный или имплицитный вопрос, и тем самым как бы толкает испытуемого реализовать в этой ситуации те или иные черты, особенности своего речевого поведения. В отличие от других лингвистических направлений в психолингвистике эксперимент становится если не ведущим, то одним из ведущих методов исследования (ср. замечание А. А. Леонтьева о том, что психолингвистика — это и есть по существу экспериментальная лингвистика).

Опора на психолингвистические модели речевой деятельности позволяет разрабатывать достаточно сложные и тонкие экспериментальные процедуры, получать и осмысливать результаты, резко изменяющие наши представления об «устройстве» механизмов речемыслительной деятельности и самого языка.

Современная психолингвистика располагает огромным арсеналом экспериментальных методик, позволяющих исследовать звуковую и смысловую стороны языка, структуру отдельного слова и группы слов, структуру отдельного предложения и целого текста. Это зачастую известные психологии виды экспериментов: на детекцию (обнаружение речевого сигнала), различение, опознание, идентификацию, интерпретацию и т. п. По большей части такие методы применялись и ранее, например в фонетике, в особенности в экспериментальной фонетике щербовской школы.

Так, уже традиционными стали эксперименты, проводимые со следующими видами инструкций:

— «Есть ли отклонения от нормы в произношении слов (в стилистическом оформлении предложений и т. п.)...?»

— «Одинаково или нет звучат гласные (интонационные контуры и т. п.)...?»

— «Запишите все, что услышите на магнитной ленте, на бланки».

— «Одинаковы ли по смыслу фразы...?»

— «Что значат слова (словосочетания)...?»

В последнее время в числе используемых психофизиологических методик можно назвать *ABX*-сравнение, при котором даются два эталонных (*A* и *B*) и один исследуемый (*X*) стимул. Испытуемые должны указать, на который из стимулов больше похож исследуемый.

В то же время сама психолингвистика если и не придумала, то значительно развила некоторые методики. К ним, в первую очередь, относятся ассоциативные эксперименты, эксперименты по шкалированию, вероятностному прогнозированию, классификациям. Ряд экспериментальных методик развился в связи с оценкой психологической адекватности трансформационной грамматики. В последнее время получила развитие методика индексирования (выделения ключевых слов) в психолингвистике текста.

Остановимся подробнее на некоторых из этих методик.

2.1. Свободный ассоциативный эксперимент. Он является базисным для исследования отдельного слова или группы слов.

Хотя начало изучения ассоциативных структур восходит еще к Аристотелю (который выделял, в частности, ассоциации по сходству, противопоставлению, пространственной и временной смежности), серьезное научное освоение результатов массового ассоциативного эксперимента стало возможным лишь в нашем столетии, после появления ряда словарей ассоциативных норм (первый такой словарь появился в 1910 г. — «Словарь ассоциативных норм английского языка» Кента — Розанова) и классических исследований Дж. Диза (1963), которому удалось показать принципиальное единство природы лингвистических (семантических) и ассоциативных группировок слов.

Методика проведения ассоциативного эксперимента чрезвычайно проста: испытуемому предлагается (устно или письменно) список слов-стимулов. После предъявления каждого слова испытуемый должен не задумываясь сказать (написать) любое пришедшее ему на ум слово-ответ (реакцию). Достоинство свободного ассоциативного эксперимента — не только в простоте его проведения, но и в легкости получения массового экспериментального материала, т. е. слов-реакций, которые поддаются многоаспектной количественной и качественной обработке.

В качестве примера рассмотрим словарную статью на слово-стимул *друг* из «Словаря ассоциативных норм русского языка». Число испытуемых — 209 человек. В статье дан самый простой вид обработки результатов эксперимента — список слов-реакций по убывающей частоте их использования в ответах испытуемых: друг — *товарищ* 39, *враг* 30, *верный* 29, *хороший* 16, *мой* 10, *недруг* 9, *близкий* 5, *настоящий, старый* 4, *брат, дорогой, надежный, преданный, приятель* 3, *закадычный, лучший, любимый, он,*

собака 2, *большой, вечный, в нужде, волк, давний, далекий, девушка, детства, добрый, дорога, друга, единственный, желанный, женщина, любовь, мальчик, милый, Мишка, муж, навсегда, не верится, нет, общий, откровенный, парта, первый, плохой, подлость, подруга, предатель, приходить, противник, сердечный, сестра, синий, собака — друг человека, честный, чудеса* 1.

В итоге получается некоторое представление «ассоциативного поля» слова *друг*. В этом поле достаточно четко выделяется его более или менее стандартное «ядро» (в нашем примере — это слова *товарищ, враг, верный*). Это ядро (около 47% всех ответов) характеризует прежде всего ассоциативные нормы, т. е. то, что именно такие слова-реакции имеют наиболее прочные ассоциативные связи со словом-стимулом. Они чаще всего будут встречаться в ассоциативном эксперименте либо для достаточно представительной группы носителей данного языка вообще, либо для людей определенного типа (если группа испытуемых сформирована по какому-то определенному признаку, например, студенты, жители определенного региона, люди определенного возраста, пола и т. д.).

Далее выделяется некоторая переходная зона слов-реакций (*хороший, мой, недруг, старый* и т. д.). Слова из этой зоны также достаточно просто связываются со словом-стимулом *друг*, однако прочность этих связей проявляется в эксперименте значительно слабее, поскольку они «забиваются» более ядерными ассоциатами.

И наконец, остальную часть словарной статьи составляет безбрежная периферия одиночных слов-реакций, которые появляются с такой частотой потому, что либо совсем подавлены ядерными и «переходными» ассоциатами, либо вообще возникли случайно, под влиянием каких-то сиюминутных ассоциаций (*в нужде, любовь, Мишка, не верится, синий* и т. д.). И если при повторном эксперименте с другими испытуемыми такой же группы слова-реакции ядра и переходной зоны с большой вероятностью должны появиться снова, то значительная часть слов-реакций периферии будет иной.

После знакомства с результатами ассоциативного эксперимента сразу же возникает проблема: как соотнести эту массу необычайно пестрого по своей структуре материала с традиционными лингвистическими представлениями и категориями. Пожалуй, наиболее распространенная сегодня классификация выделяет три вида ассоциаций: парадигматические, синтагматические и тематические.

В парадигматических ассоциациях слово-реакция выступает как соотносимое со словом-стимулом либо «по горизонтали» (отношение «координации»: слова-синонимы *друг — товарищ* или антонимы *друг — враг*), либо «по вертикали» (отношение «субординации», род — вид: *фрукт — яблоко, друг — Мишка,* —

или отношение «суперординации», вид — род: *стул — мебель, друг — он*).

В синтагматических ассоциациях слова-реакции дополняют слово-стимул до «нормального» словосочетания (*друг — приходить, друг — верный*).

Наконец, третий вид ассоциаций — это так называемые тематические ассоциации, когда слово-реакция может употребляться вместе со словом-стимулом в рамках тематически ограниченного контекста (предложения): *друг — любовь, друг — чудеса, друг — не верится, друг — парта.* Объясняя механизм тематических ассоциаций, некоторые исследователи считают, что парадигматические и синтагматические ассоциации являются «собственно лингвистическими» и основаны на языковой природе слова-стимула (парадигматические — на системных отношениях с другими единицами, синтагматические — на отношениях с контекстуальным окружением), а тематические ассоциации являются «экстралингвистическими», так как основаны на отражении в сознании взаимодействия денотата (сигнификата слова-стимула) с другими объектами реального мира.

Разумеется, возможны и другие аспекты анализа ассоциаций (например, выделение деривационных ассоциаций к корню слова: *друг — друга*; выявление основных типов ассоциаций для разных частей речи и т. д.).

При обработке ассоциаций исследователи используют различные коэффициенты, например, отношение числа разных реакций к числу всех реакций. Можно предложить в качестве оценки компактности распределения и оценку энтропии H или же величину 1/H, которую можно трактовать как «ассоциативную силу» слова (Овчинникова, Штерн, 1989).

Собственно для лингвистики результаты свободного ассоциативного эксперимента показывают недостаточность слишком «логизированных» описаний структуры лексикона и заставляют видеть реальные сложности и противоречия в формировании структур связей.

Стремление уйти от «засилья» самых частых слов-реакций, мешающих более «объемным» исследованиям ассоциативных структур, вызвало к жизни огромное число модификаций свободного ассоциативного эксперимента. Так появился направленный ассоциативный эксперимент, некоторым образом вводящий ограничения либо в сам стимул, либо в экспериментальное задание. Самый простой вид направленного ассоциативного эксперимента — это ограничение класса слов-реакций, ассоциатов (например, задание написать не вообще любое слово, а, например, любое прилагательное).

2.2. Методика семантического дифференциала Ч. Осгуда. Она возникла как оценка стимула в каких-либо признаках. Поэтому она, с одной стороны, основана на ассоциации стимула с признаком, а с другой стороны, мо-

жет рассматриваться как шкалирование, ибо для каждого признака (шкалы) задается степень силы этой связи. Например:

:3:2:1:2:3:

добрый	злой
высокий	низкий
слабый	сильный
красивый	некрасивый
пассивный	активный

и т. п.

Заметим, что обычно используются семибалльные шкалы, хотя возможны 5- или 3-балльные.

В задачу испытуемого входит соотнести предъявленные стимулы с этими шкалами, проставив оценку, т. е. тесноту связи. Например, предъявляется слово *дедушка*, и надо его оценить в шкале «добрый — злой». Обычно большинство испытуемых оценивает это слово как «добрый» с баллами 2 или 3. Для чего нужны эти оценки? Они позволяют для каждого стимула получить набор шкал с числовыми оценками. Эти наборы позволяют построить поле признаков, в котором описывается понятие, обозначенное словом, т. е., по сути, поле направленных ассоциаций. Эти поля дают возможность сравнить разные слова одного языка, например, близкие по значению, с тем, чтобы найти между ними различие, или же сравнить эквиваленты из разных языков.

При этом можно пользоваться как индивидуальными оценками, так и усредненными по группе испытуемых. В последнем случае в качестве показателя центральной тенденции корректнее брать не среднюю арифметическую, а медиану.

Методика Ч. Осгуда получила большое распространение в психолингвистических исследованиях. Например, были получены оценки для звуков языка; они позволили найти степень близости между звуками, позволили сравнить оценки у билингвов, сравнить оценки корреспондирующих звуков из разных языков. Методика шкальных оценок с успехом применяется и для оценки целых текстов (радиорепортажей, научно-популярных и даже художественных текстов).

2.3. Вероятностное прогнозирование. Вероятностное прогнозирование как психолингвистическая методика включает в себя по крайней мере два момента: 1) определение субъективных частот речевых элементов с помощью шкалирования, 2) использование имплицитного «знания» об этих частотах в экспериментах или в естественной коммуникации (Р. М. Фрумкина).

Деятельность человека вообще опирается на вероятностный прогноз, т. е. человек обладает способностью использовать накопленный индивидуальный опыт для прогнозирования ситуации и наибольшей эффективности действий. Все это в полной мере может быть отнесено и к речевой деятельности. Исходная ги-

потеза состоит в том, что единицы высказывания (звуки, буквы, слова, слоги и т. п.) «записаны» в мозгу человека с частотами появления их в речи. Эти субъективные частоты информант способен оценить. Так, если предложить информанту набор слов и шкалу частотности, то он легко распределяет эти слова по градациям, например, от «никогда не встречается» до «встречается на каждом шагу». В сущности, эта процедура мало отличается от оценочной методики Осгуда.

В результате обработки (а в этом случае используется аппарат непараметрической статистики) можно получить усредненные оценки $F_{суб}$, согласно которым можно построить словари субъективных частот. При сравнении их со словарями объективных частот вскрываются две интересные закономерности: а) слова так называемой бытовой лексики (*иголки, ботинки* и пр.) получают завышенные оценки, т. е. оцениваются в данном случае не слова, а реалии; б) наибольшую вариативность в группе испытуемых получают профессионализмы.

Метод субъективного шкалирования может использоваться в клинике как диагностирующий тест. Так, при шизофрении больные дают оценки, существенно отличающиеся от полученных в норме.

Исследование влияния вероятности стимулов (и неречевых, и речевых) на их восприятие проводилось в ряде психолингвистических экспериментов. Так, показано, что чем больше вероятность, например, слова, тем ниже пороги его распознавания в условиях помех (т. е. допускается больший уровень помехи). Это верно и для зрительного, и для слухового восприятия. В общем виде эта зависимость имеет логарифмический характер. Можно также утверждать, что чем частотнее слово, тем больше процент его правильного восприятия группой испытуемых (или тем лучше оно запоминается и т. п.). Важно отметить, что эти зависимости сильнее, если используются не $F_{об}$, а $F_{суб}$.

2.4. Индексирование текста. Индексированием называется выделение ключевых (основных смысловых) слов текста. Можно предположить экспериментально-статистическую методику, при которой индексирование проводится группой испытуемых, далее все выделенные слова получают частоту их появления (т. е. индексирования) в эксперименте. В результате появляется список слов с их частотами, внешне похожий на статью словаря ассоциативных норм. В этом списке есть «ядро», которое отражает общее понимание текста, и единичные индексации, отражающие индивидуальное понимание текста. В высокочастотной зоне можно выделить три набора ключевых слов: малый (обычно 1—4 самых частых слова), средний (первые семь — десять слов) и большой (около 15—20 слов). Эти три набора, как правило, последовательно отражают основную тему, основное взаимодействие действующих лиц (конфликт) и основное содержание текста. С помощью на-

бора ключевых слов можно исследовать, с одной стороны, структуру плана содержания текста, сравнивать с этой точки зрения тексты разных жанров, устные и письменные версии и т. п., а с другой стороны, понимание текста разными людьми, разными социальными группами и пр.

3. **Особенности материала психолингвистических исследований.** Как уже отмечалось. характерной чертой психолингвистики является не просто учет, но активная ориентация на «нестандартные», «ненормативные» виды коммуникации. Это детская речь, речь при различного рода патологиях, при коммуникации в затрудненных условиях, на иностранном языке (при плохом знании его), речь в состоянии эмоционального возбуждения (даже стресса), коммуникация при помехах в канале связи или в искусственных человеко-компьютерных системах (например, в современных автоматизированных ИПС). Наконец, по-новому видятся в этом аспекте и общие закономерности разговорной речи, которая может быть рассмотрена не как «отклонение» от кодифицированного литературного языка, а как один из видов реализации коммуникативных структур в особых условиях.

Материал исследования часто разделяют на «нормальный» и «патологический», т. е. полученный от нормальных взрослых испытуемых и от больных с разного рода центральными патологиями. Однако многие феномены не укладываются в эту классификацию (прежде всего — речь в затрудненных условиях, детская речь и речь на иностранном языке). Это заставляет предложить другую классификацию материала: 1) речь без помех, 2) речь в условиях помех.

В данном разделе мы остановимся прежде всего на речи в условиях помех как менее известной лингвистам.

Речь в условиях помех, по-видимому, весьма распространена в естественной коммуникации; при этом сами помехи могут быть разных типов. Это, прежде всего, объективные помехи. К ним относятся различные виды шумов (в исследованиях часто используется экспериментальный белый шум, т. е. ровный шум в речевой полосе от 50 до 5000 Гц), прием речи с большого расстояния, кратковременное предъявление письменной речи с экрана. Сюда же можно отнести результаты восприятия речи при различных модификациях сигнала — синтезированной или фильтрованной речи, речи, искаженной акцентом, и т. п. Речь, искаженная объективной помехой, возникает и при недостатках речевого тракта говорящего (короткий язык, «волчья пасть») и логопедических нарушениях (картавость, шепелявость и т. п.). Субъективные помехи связаны с «центральными» патологиями или отклонениями в глубинных механизмах речепроизводства и речевосприятия. Так, с одной стороны, это речь больных с афазией, шизофренией и другими заболеваниями. С другой стороны, это детская речь или речь на иностранном языке.

Речь в помехах получила в психолингвистике такое широкое распространение потому, что она позволяет найти глубинные структуры речевой деятельности. И. П. Павлов писал: «Патологическое часто открывает нам, разлагая и упрощая, то, что заслонено от нас, слитое и усложненное, в физиологической норме» (Павлов, 1951, с. 317). Результаты ряда экспериментов показывают, что это положение можно распространить не только на собственно патологию речи, но и на все другие виды помех.

На некоторых видах материала психолингвистических исследований остановимся более подробно.

3.1. Детская речь. Она дает в руки исследователя уникальный материал: с одной стороны, она показывает процесс овладения языком, а с другой стороны, помогает понять язык, речевую деятельность взрослого человека.

Традиционное обращение лингвистов к фактам детской речи обычно ограничивается наиболее яркими отличиями в речи детей от «взрослой речи» (типа *почтанник* вместо *почтальон* или *сыски* вместо *шишки*). Не случайно ими занимались скорее писатели (например, К. И. Чуковский, В. Драгунский), чем ученые-лингвисты.

Психологов речи «тайна» процесса овладения языком занимает уже несколько столетий. В наше время этот феномен — в центре внимания психолингвистики. Имевшая определенное распространение теория о том, что ребенок овладевает речью через имитацию и употребление в различных ситуациях готовых речений, не может объяснить, как же в речи появляются те формы, которых ребенок ранее не слышал. Ясно, что ребенок переходит к постижению правил пользования элементами языка — от фонологических до синтаксических. Кроме чисто научного интереса есть и большое прикладное значение этой проблемы — как для клиники детских речевых патологий, так и для проблемы обучения языку (и родному, и иностранному). В настоящее время признано, что на каждом этапе речевого развития ребенок владеет сложившейся у него к тому времени системой, и в течение нескольких лет жизни происходит ряд смен таких систем. Становление речи идет с самого раннего возраста и продолжается вплоть до подросткового.

Первые этапы речевого проявления — гуление и лепет. Это еще не речь, но в них заложены предпосылки усвоения фонетики определенного языка, предпосылки становления артикуляционной базы. На этапе гуления (около 3,5 мес.) в речи ребенка могут появляться все возможные артикуляции, звуки всех языков мира. В процессе общения со взрослыми ребенок слышит лишь определенные звуки и на этапе лепета (около 4—8 мес.) подражает им, а затем отбирает и формирует те противопоставления, которые оказываются важными для коммуникации, для общения со взрослыми. При этом вначале он

учится различать и лишь потом воспроизводить звуки в их основных системных противопоставлениях (в этой связи психолингвисты говорят о примате перцептивной базы перед артикуляторной).

Становление осмысленной речи идет вслед за усвоением ребенком предметной деятельности во всех ее проявлениях — игровой, прагматической, познавательной. Первые «слова» (около 9—12 мес.) — чаще всего звукоподражательные комплексы, которые означают и ситуацию, и предмет одновременно. Постепенно усваиваются и реальные слова языка. Они свидетельствуют о вычленении предметов из ситуации и начинают функционировать в роли слова-предложения, будучи внутренне предикативными (это уже приблизительно в год). Употребляя их, ребенок лишь называет основную информацию, взрослые же понимают их только в конкретной ситуации. После того, как ситуация расчленяется (приблизительно в 1,5 года), в речи ребенка появляются двусоставные предложения типа «Мама ту-ту-у», где и тема и рема получают свое выражение.

Вторая половина второго года жизни ребенка — это период «открытия» новых слов. С этого времени и вплоть до школы слово становится неотъемлемым признаком предмета, его принадлежностью. Поэтому в преддошкольном и дошкольном возрасте дети не справляются с экспериментом по переименованию предметов. Так, даже в игровой ситуации они отказываются корову назвать лошадью и наоборот. На первых порах дети не считают словами глаголы и — тем более — служебные части речи. Например, на инструкцию сказать, какие слова есть в предложении *Книга лежит на столе,* отвечают: «*Книга* и *на столе*».

С развитием механизма расчленения ситуации и осознания участников ситуации и их ролей развивается и механизм вычленения элементов речи и их комбинаций. Это связано, с одной стороны, с усвоением синтаксических отношений, а с другой стороны, с выделением морфем и осознанием словообразовательных моделей. Начинается известный возраст «словотворчества» (от 3 до 5 лет).

Первые самостоятельные «взрослые» предложения ребенка имеют структуру «субъект — предикат — объект». Части речи, которые наиболее распространены, — существительные и глагол. В собственной детской речи очень мало прилагательных. А предлоги и союзы, хотя и употребляются в речи, осознаются как особые слова языка лишь после обучения в школе.

Широко распространено мнение, что к 7 годам языковая система в основном сформировалась, точнее — она такая же, как у взрослых. Однако это не совсем так. Еще в течение нескольких лет будет стремительно расти словарный запас. Следует отметить, что это не механическое увеличение объема лексикона. Как показали психолингвистические исследования, лек-

сикон в сознании человека организован особым образом: это сложная структура, состоящая из семантических полей, синонимических и антонимических отношений; кроме того, слова «записаны» в памяти человека с их частотами употребления в речи. Опыты на вербальные ассоциации показывают, как с 1-го по 10-й класс меняется их структура: если в начале преобладают предметные ассоциации, то к концу школы формируется такое же соотношение различных видов ассоциаций, как у взрослых испытуемых. Кроме того, в первых классах встречается (и довольно много) ассоциаций по звуковому составу. А умственно отсталые дети и в более старшем возрасте дают в основном звуковые ассоциации, т. е. у них остается еще тот же интерес к звуковой оболочке, что у маленьких детей.

Сложные предложения появляются в речи детей в дошкольном возрасте. В то же время последние исследования говорят о том, что еще и в подростковом возрасте ребенок не в полной мере владеет самым высшим — текстовым — уровнем речи. Поскольку в онтогенезе диалогическая речь предшествует монологической, построение цельного, развернутого, монологического текста представляет для ребенка значительные трудности.

Таким образом, изучая детскую речь, мы имеем возможность проследить, как в онтогенезе формируется языковая способность и развивается языковая активность. Знание механизмов важно в прикладной работе — когда лингвисты помогают составить обучающие программы для детей с задержкой речевого развития.

3.2. Патология речи. Изучение структур мозга является в настоящее время очень актуальной темой. И в нашей стране, и за рубежом ведутся работы по изучению специализации различных участков мозга в деятельности человека. И один из важнейших вопросов — как и где в структурах мозга записывается специфическая речевая информация.

Материалы разного рода патологии речи гораздо менее известны лингвистам, чем детская речь. К тому же синдромы здесь порой не очень четко различаются по проявлениям в речи. При этом получение и — особенно — обработка материалов во многих случаях резко затруднены, материалы чаще всего очень плохо поддаются исследованию с помощью традиционных лингвистических описаний, «грамматик». Все это затрудняет использование такого материала лингвистами. Но в то же время данные эти для выявления общих закономерностей речевой деятельности бесценны. Не случайно известная модель порождения высказывания, разработанная в конце 60-х годов А. А. Леонтьевым и Т. В. Рябовой (Ахутиной), в значительной мере базировалась на данных патологии речи.

3.2.1. Патология речи при афазиях. Афазиями называются расстройства речи при органических поражениях мозга. Такие нарушения имеют внутренний характер, т. е. не связаны

с повреждениями речевого аппарата, а возникают вследствие очагового поражения коры головного мозга при травмах или ранениях, а также опухолевых и сосудистых заболеваниях.

А. Р. Лурия отмечает существование двух раздельных систем коры головного мозга, одна из которых связана с работой задних (теменно-затылочно-височных) отделов и обеспечивает прием, переработку и хранение информации, а вторая, связанная с работой передних (премоторных и лобных) отделов, ведает формированием намерений, планов и программ поведения. Задние отделы левого полушария коры обеспечивают кодирование информации в парадигматические ряды языка. Поражение этих участков приводит к нарушениям в системах языковых кодов, т. е. к нарушениям парадигматических отношений, причем синтагматические отношения сохранены. Обратное наблюдается при поражениях передних отделов речевых зон: парадигматические отношения практически сохраняются, но грубо нарушена развернутая речь, иначе говоря, синтагматические отношения. Таким образом, можно утверждать, что при речевой деятельности парадигматическая и синтагматическая организации речевого процесса обеспечиваются разными системами мозга. Именно это положение лежит в основе выделения двух основных классов афазических нарушений речи, связанных с расстройством или парадигматических, или синтагматических отношений.

Обычно различаются следующие группы афазий: височная (сенсорная) и затылочно-теменная (семантическая), возникающие при поражении задних отделов, и лобная (моторная и динамическая), связанная с поражением передних отделов.

При сенсорной афазии, описанной немецким ученым К. Вернике во второй половине прошлого века, больной перестает понимать чужую речь, воспринимает речевые звуки как нечленораздельные шумы, ему трудно называть предметы, повторять слова, писать под диктовку. В современных классификациях выделяются разные виды сенсорной афазии. У больного с акустико-гностической афазией (классическая форма) нарушен фонематический слух. Это расстройство выражается в трудностях дифференцировки фонем. Одно и то же слово звучит для больного каждый раз иначе, хотя и сохраняет некоторый общий мелодический рисунок. При повторении это приводит к изменениям звукового состава слова (литеральным парафазиям — ошибочным заменам элементов речи в рамках парадигматического класса), а также к заменам нужного слова другими, близкими по звучанию или смыслу (вербальным парафазиям). Нарушаются письмо и чтение. Больше всего страдают слова с конкретным, вещественным значением. Грамматическое значение слова обычно сохраняется: нестойкие в звуковом отношении слова, как правило, не утрачивают своего грамматического значения. Больные говорят много и быстро, встречаются в изобилии междометия, вводные слова, предлоги, местоимения,

глаголы, а существительных сравнительно мало, и в предложениях нередко отсутствуют подлежащие. Интонационная и мелодическая структура речи сохраняется. В тяжелых случаях речь больного превращается в так называемый «словесный салат»: он говорит быстро и легко, но понять, что он хочет сказать, трудно. Непостоянное в звуковом отношении слово теряет свою предметную отнесенность, что приводит к явлению, называемому «отчуждением смысла слов».

При акустико-мнестической афазии фонематический слух сохраняется, основной дефект — мнестический: больной забывает названия предметов. Ему трудно запомнить и воспроизвести серию из нескольких слов, особенно в усложненных условиях, например, с интервалом между словами. Трудности возрастают при повторении предложений. Речь больного свободна, оформлена интонационно, но изобилует вербальными парафазиями.

При семантической афазии, впервые выделенной в первой половине XX в. Х. Хэдом (Кембридж), не наблюдается ни нарушений фонематического слуха, ни нарушений артикуляций. Основные дефекты заключаются в расстройстве словесной памяти и в трудностях понимания сложных логико-грамматических структур. Так, больные не справляются с пространственными и временными конструкциями с предлогами *под*, *над*, *перед*, *после*, с атрибутивным родительным падежом, обратным порядком слов, сравнительными конструкциями, страдательным залогом. Например, больной не может сказать, какое предложение правильно: *Солнце освещается землей* или *Земля освещается солнцем*. Расстройство речевой деятельности при семантической афазии внутренне связано с нарушением ориентировки в пространстве.

Моторная афазия была впервые описана французским ученым П. Брока в 1861 г. В настоящее время различается несколько видов моторной афазии. Больной может лишиться способности свободно, без напряжения артикулировать звуки речи или слова. В таком случае мы имеем расстройство техники звукопроизводства, чисто артикуляторные нарушения. Это афферентная моторная афазия. Характерный дефект речи такого рода больных — замены артикуляций, причем, как правило, замещают друг друга фонемы, противопоставленные по одному дифференциальному признаку. При тяжелых формах могут смешиваться и артикуляторно далекие фонемы, различающиеся несколькими дифференциальными признаками. Больные пишут с грубыми ошибками и характерными артикуляторными заменами. Но переключение с одной артикуляции на другую происходит у больных значительно легче. Они относительно легко произносят целые фразы, сохраняется мелодическая структура речи.

При эфферентной моторной афазии наблюдаются моторные нарушения другого вида. Больным трудно переходить от пре-

дыдущей артикуляции к последующим, хотя способность артикулировать отдельные звуки часто сохраняется; происходит как бы «застревание» на каком-либо звуке или слоге. Если сохраняются слова, больным трудно построить предложения из слов. Разрушается мелодика речи. Аналогичным образом нарушается и письмо, например, больной может написать отдельные буквы, но ему трудно писать слоги или слова.

Наиболее поздняя стадия обратного развития эфферентной моторной афазии может принимать форму нарушения спонтанной речи, когда больной не в состоянии самостоятельно построить развернутое высказывание. Такие нарушения встречаются и как отдельная форма речевых расстройств. В работах А. Р. Лурия и его учеников это расстройство называется динамической афазией.

Выдающийся отечественный психолог Л. С. Выготский в своей работе «Мышление и речь» (1934) описал характеристики внутренней речи и обратил внимание на ее своеобразный предикативный характер. Некоторые исследователи, отмечая, вслед за Л. С. Выготским, предикативный характер внутренней речи, говорят об ее относительной сохранности при сенсорной афазии. Для сенсорной афазии характерен именно предикативный строй речи, сохраняется мелодика речи. В то же время отмечается, что при эфферентной моторной и динамической афазии внутренняя речь нарушается или, точнее, нарушены какие-то внутренние этапы порождения высказываний.

А. Р. Лурия писал о распаде внутренней синтаксической схемы высказывания у больных с нарушенной предикативной функцией речи, что во внешней речи выражается в феномене так называемого «телеграфного стиля». Он наиболее четко проявляется на сравнительно поздних стадиях обратного развития эфферентной моторной афазии. Речь таких больных изобилует самыми грубыми грамматическими ошибками, хотя встречаются грамматически правильные стереотипные обороты. Служебные слова опускаются, преобладают существительные, в основном в именительном падеже. Замедленный темп речи, напряженность, утомляемость, остатки аграмматизма могут наблюдаться у больных годами.

Известный лингвист Р. О. Якобсон выдвинул интересную гипотезу, которая сводится к следующему.

Процесс распада речи при афазии зеркален процессу становления речи в онтогенезе, а восстановление ее у больных повторяет процесс становления речи у ребенка. Это значит, что более сложные формы, такие как сложные синтаксические образования, редкие слова, более «трудные» звуки речи, при нарушении деятельности левого полушария пропадают прежде всего. Сохранным остается в основном то, что было усвоено раньше на каждом из этапов речевого развития.

Эта гипотеза получила ряд экспериментальных подтверж-

дений (хотя в то же время имеются данные, противоречащие ей).

Исследования речи при различных субъективных и объективных помехах позволяют расширить гипотезу Р. О. Якобсона следующим образом: при любого рода помехах более сохранно то, что усвоено раньше в онтогенезе.

3.2.2. Патология речи при угнетении одного из полушарий головного мозга. Известно, что афазии возникают в основном при поражениях левого, ведущего для речи доминантного полушария. Правое полушарие до недавнего времени считалось не имеющим отношения к речи. Однако в настоящее время это мнение требует пересмотра.

Сравнительно новые факты по асимметрии полушарий головного мозга во многом революционно изменили наши представления о речемыслительной деятельности человека. Эти данные получены, в частности, при лечении различных видов депрессии с помощью электрошоков. Больной получает электрический разряд через электроды, приставленные к одной половине головы. После судорожного припадка в процессе восстановления нормальной работы мозга наступает такой период, когда одно полушарие еще угнетено («выключено»), а второе уже работает (иногда даже более активно, чем в норме). В это время (оно длится до 40 минут) ведутся клинические наблюдения, а также психологическое и психолингвистическое тестирование. В отношении речевой деятельности в настоящее время можно считать установленной следующую картину специализации полушарий.

При угнетении левого полушария, которое является у правшей ведущим для речи («доминантным»), речевая деятельность резко отличается от нормальной: страдает как порождение, так и восприятие речи. Афазии, возникающие при левосторонних припадках, вначале идут по типу тотальной, а по мере восстановления функций левого полушария — по смешанному сенсомоторному типу. Исследования фонематического слуха говорят о том, что нарушен переход от восприятия тональных сигналов к фонемным обобщениям, замены при восприятии согласных идут в основном по одному из дифференциальных признаков. Но еще более, чем восприятие звуков языка, страдает слухоречевая память на слова. В собственной речи больных значительно уменьшается объем высказываний, резко упрощается синтаксис и увеличивается доля конкретных существительных и прилагательных, т. е. лексика становится более предметной и оценочной. Более предметными становятся ответы в ассоциативном эксперименте. Больные при «выключенном» левом полушарии не могут проводить синтаксические трансформации, владеют лишь простейшим, «глубинным» синтаксисом. При этом у билингвов синтаксирование на втором языке нарушается на всех уровнях: они могут понять речь на втором языке, но при

ответах непременно переходят на свой первый язык. Из механизмов высших уровней страдает также вероятностное прогнозирование. В то же время при «освобождении» от действия левого полушария обостряется слух, в частности музыкальный. Кроме того, опознание речевых отрезков в условиях помех обеспечивается даже лучше, чем в норме.

При инактивации (угнетении) правого полушария классическая картина афазии практически не наблюдается. Однако правое полушарие оказывается не безразличным к речевой деятельности. Прежде всего замечено, что при правосторонних припадках наблюдаются расстройства просодики, выражающиеся как в переносе ударения, так и в монотонности речи, неадекватном восприятии интонационных контуров, в частности эмоциональной мелодики. С общим расстройством тонального слуха связано и неразличение мужских и женских голосов, неузнавание знакомых по голосам. Восприятие речи в условиях помех резко затруднено. В собственной речи больных наблюдается многоречивость, так как правое полушарие «не сдерживает» активности левого. Увеличивается доля глаголов и служебных частей речи. При резком увеличении числа морфологических дериватов и сложном внешнем синтаксисе больной зачастую не понимает основной темы в обращенной к нему речи, даже очень простой по внешнему оформлению. Предположительно это может быть связано с нарушением глубинных синтаксических отношений. В ассоциативном эксперименте в ответах появляются отвлеченные существительные — «синтаксические дериваты» (с суффиксами *-ние, -ость* и т. п.) и слова, однокорневые со словом-стимулом.

Таким образом, левое полушарие отвечает за абстрактно-логическое построение мысли. В нем локализованы сложные механизмы для аналитической обработки информации на разных лингвистических уровнях — от схемы синтаксических трансформаций до системы дифференциальных признаков фонем.

Правое же полушарие отвечает за непосредственную, конкретную, моментальную связь с предметным миром, за глобальное освоение ситуации. Осуществляя денотативную связь, оно является в основном хранилищем конкретных существительных. В правом полушарии локализованы механизмы просодической организации высказывания (и, следовательно, механизмы актуального членения речи). Осознание текста на глубинном, содержательном уровне («цельность» текста) происходит также при работе правого полушария. Правое полушарие помехоустойчиво. И еще одна очень важная его функция — регулировать (тормозить) активность левого полушария.

Полученные данные об особенностях речи при афазиях и унилатеральных припадках позволяют не только в принципе «локализовать» в мозгу те или иные процессы, механизмы, но и показывают реальную относительную самостоятельность неко-

торых «узлов» нашего речевого механизма, а также позволяют оценить адекватность наших лингвистических (точнее — логико-лингвистических) построений тем речемыслительным механизмам, которые «вырисовываются» на основе данных патологии речи.

3.2.3. Патология речи при шизофрениях и маниакально-депрессивном психозе. Если при афазии основное нарушение — это нарушение речевого механизма при относительной сохранности мышления, то при шизофрениях разных форм и при маниакально-депрессивном психозе ведущим является нарушение мышления или эмоциональный сферы, что выражается в нарушении коммуникативной способности. Это нарушение в общем виде, по мнению большинства психиатров, связано с нарушением механизма вероятностного прогнозирования, что, в свою очередь, возможно, вызывается нарушением функциональной асимметрии полушарий головного мозга.

Первое впечатление, которое вызывает речь шизофреника, это (при легких и средних формах) резонерство, невнятность, бессвязность, переходящая при тяжелых формах в «словесный салат», т. е. цепочку слов, их обрывков или словосочетаний, отдельных выкрикиваемых звуков. На фонетическом уровне, кроме нарушений четкости артикуляции, иногда — громкости и темпа, специфических изменений не обнаруживается. В то же время сами больные проявляют повышенное внимание к звукам речи, аллитерациям, рифме и придают им символическое значение, а при письме демонстрируют вычурное написание букв, украшательство. Основные же отклонения связаны с высшими лингвистическими уровнями.

Прежде всего нарушается организация текста, самый высший уровень речевой деятельности; это выражается в дефектах как цельности, так и связности текста. Например, при задании вставить в текст пропущенные слова больной совершает ряд ошибок: «Один человек заказал пряхе тонкие *носки* (вместо *нитки*). Пряха спряла тонкие нитки, но человек сказал, что нитки *тонкие* (вместо *толстые*) и что ему нужны нитки самые тонкие. Пряха сказала: „Если тебе эти не тонки, то вот тебе другие", — и она показала на *нитки* (вместо *пустое*) место. Он сказал, что не видит. „Оттого и не видишь, что они очень тонкие, я и сама их не вижу"».

По мнению большинства исследователей, в речи больных проявляются прежде всего нарушения феномена коммуникации, заключающиеся в том, что больной не заботится о том, чтобы его понимали. Такая речь «для себя», а не «для других» является проявлением аутизма, т. е. отгороженности от мира, сосредоточенности на своем внутреннем мире.

При исследовании синтаксиса отмечается, что в речи шизофреников меньше дополнений и обстоятельств, а при маниакально-депрессивных психозах предложения могут редуциро-

ваться до нераспространенных. Хотя грамматически речь в основном правильна (при не самых тяжелых состояниях), встречаются нарушения управлений. Что касается лексики, то нарушение вероятностного прогноза ведет к неправильному выбору из синонимического ряда. Так, в нейтральной речи появляются *очи* вместо *глаза*, различные сложные терминологические сочетания и т. п. Например, на вопрос, что такое *шкаф*, больной отвечает: «Шкаф — это вещь, относящаяся к *неживой природе*; она имеет *прикладное* значение для сохранения других *материальных частиц*». Выявленные в эксперименте нарушения классификации слов больными могут, по-видимому, свидетельствовать как об искажении стандартов на уровне обобщения, так и о неустойчивости структурной организации лексикона. Например, больной объединяет слова *шкаф* и *кастрюля* в одну группу потому, что «у обоих предметов есть отверстие». А при задании сравнить понятия «сани» и «телега» отвечает: «Оба эти слова с точки зрения грамматики являются существительными, но сани не бывают в единственном числе».

Зачастую больные конструируют новые слова — по правилам языка или изобретая собственные словообразовательные модели. В ассоциативных экспериментах больные дают в целом нестандартные ассоциации.

Таким образом, первичный дефект при данном заболевании — не специфически речевой. По модели порождения речи Леонтьева — Рябовой, он затрагивает лишь самые глубинные уровни. Однако нарушения мышления приводят и к нарушению в речи. По-видимому, ближайшие задачи для психолингвистики в этом плане — разобраться, каковы реальные источники этих нарушений и какие системные речевые нарушения выявляют особенности дефекта на глубинных уровнях. Отметим здесь только, что у специалистов имеются серьезные основания полагать, что феномен шизофрении связан с угнетением правого полушария.

3.3. Хезитации и аграмматизмы в устной спонтанной речи. Естественно, что материал такого типа получают в основном с помощью наблюдения. Хезитация — это колебание в речи говорящего. Это может быть чистая пауза или заполненная вокализациями типа «м-м-м», «э-э», словами *вот, как это, ну*. Такие хезитации возникают тогда, когда говорящий не успел спланировать речь, когда нужно дополнительное время на подбор лексемы, синтаксической конструкции и т. п.

Под аграмматизмом понимается грамматическая ошибка в речи. В ряде случаев она может возникать из-за плохого знания языка (например: *оплатите за проезд, яичная шампунь*), оговорок (*посвящен изучению афазию*), перепланирования речевой программы (*такой большой ... фабрика*).

Парафазии (они, как уже отмечалось, характерны и для

речи больных с сенсорной афазией) — это ошибочные замены элементов в речи в рамках парадигматического класса. Так, во фразе *черно-белый холодильник* произошла непроизвольная замена слова *телевизор* на слово *холодильник* в рамках парадигматического класса, который условно можно назвать «наименования сложной бытовой техники»; это так называемая вербальная парафазия. Во фразе *я снесу вам яичко другое — не простое, а золотое* вместо *не золотое, а простое* — парафазия в путанице акцентов антонимического противопоставления.

Все эти факты служат для познания организации наших лингвистических знаний и речевого процесса, ибо, как сказал один психолог, «существуют правила для нарушения правил». Так, паузы хезитации могут помочь решению проблемы единиц порождения речи; парафазии же говорят об организации элементов (в данном случае — лексики) в семантические поля или классы похожих слов и т. п.

* *
*

Резкое расширение материала психолингвистического исследования, рассмотрение коммуникации с помощью кодифицированного литературного языка лишь как частного случая речевой деятельности (хотя и очень важного, а в некоторых видах коммуникативных ситуаций даже ведущего), широкое использование данных разнообразных психолингвистических экспериментов — все это позволяет построить более общие модели речемыслительной деятельности, чем это возможно в традиционной лингвистике, опирающейся на данные традиционного материала исследования. Психолингвистическое моделирование позволяет увидеть общие закономерности речевой деятельности, общее и особенное во всех формах коммуникации, выявить «базисные», глубинные, и более поверхностные механизмы, объяснить их роль и функционирование в механизмах речемыслительной деятельности, объяснить многие явления, связанные с реальной деятельностью в условиях разных помех, и с тем, как строится система языка, приспосабливающаяся к этим условиям, и как вообще работают механизмы построения и восприятия (осознания) текста. Вырисовывается возможность создания достаточно глобальной и в то же время адекватной динамической модели речемыслительной деятельности. Эта модель еще не создана в деталях, но некоторые ее основные контуры весьма интересны и достаточно нетрадиционны.

В заключение отметим, что не случайно раздел о психолингвистике помещен в книге о прикладном и математическом языкознании. Теория речевой деятельности имеет ярко выраженный прикладной характер. Это определяется не только и не столько методами исследования (эксперимент в настоящее время предполагает с необходимостью статистико-вероятностное

моделирование), сколько применением результатов. А они могут применяться в инженерной психологии (улучшать работу диспетчеров), в теории связи (сделать более надежной работу операторов, имеющих дело с речевой информацией в условиях помех), в методике преподавания языков (учесть специфику механизма родного и изучаемого языков, особенности взаимодействия этих механизмов), в пропаганде (использовать механизмы и особенности речевого воздействия). И, конечно, особое значение имеют результаты психолингвистических исследований в патологии речи. Придя в клинику, лингвист не только проводит тесты с больными для того, чтобы познать механизмы речи, теоретически осмыслить эти результаты, но, познав что-то, он должен принять участие в составлении диагностирующих тестов, в создании артикуляционных таблиц для слухового протезирования, в составлении упражнений и методик для восстановительной работы с больными.

Литература

Ахутина Т. В. Порождение речи: Нейролингв. анализ синтаксиса. М., 1989.

Балонов Л. Я., Деглин В. Л., Черниговская Т. В. Функциональная асимметрия мозга в организации речевой деятельности // Сенсорные системы: Сенсорные процессы в асимметрии полушарий / Под ред. Л. И. Леушиной, А. В. Бару. Л., 1985.

Волошинов В. Н. Марксизм в философии языка. М., 1929.

Выготский Л. С. Мышление и речь // Собр. соч.: В 6 т. Т. 2. М., 1982.

Горелов И. Н. Разговор с компьютером: Психолингв. аспект проблемы. М., 1976.

Жинкин Н. И. 1) Механизмы речи. М., 1958; 2) Речь как проводник информации. М., 1982.

Залевская А. А. Слово в лексиконе человека: Психолингв. исслед. Воронеж, 1990.

Зейгарник Б. В. Психопатология. М., 1986.

Иванов В. В. Чет и нечет. М., 1980.

Исследование речевого мышления в психолингвистике / Отв. ред. Е. В. Тарасов. М., 1985.

Левицкий В. В., Стернин И. А. Экспериментальные методы в семасиологии. Воронеж, 1989.

Леонтьев А. А. 1) Слово в речевой деятельности. М., 1955; 2) Психолингвистические единицы и порождение речевого высказывания. М., 1969; 3) Психолингвистика // Тенденции развития психологической науки. М., 1989.

Лурия А. Р. Язык и сознание. М., 1979.

Мурзин Л. Н., Штерн А. С. Текст и его восприятие. Свердловск, 1991.

Негневицкая Е. И., Шахнарович А. М. Язык и дети. М., 1981.

Новиков В. И. Семантика текста и ее формализация. М., 1983.

Основы теории речевой деятельности / Под ред. А. А. Леонтьева. М., 1974.

Петренко В. Ф. Психосемантика сознания. М., 1988.

Психолингвистика / Общ. ред. А. М. Шахнаровича. М., 1984.

Сахарный Л. В. 1) Психолингвистические аспекты теории словообразования. Л., 1985; 2) Введение в психолингвистику: Курс лекций. Л., 1989.

Слобин Д., Грин Дж. Психолингвистика. М., 1976.

Словарь ассоциативных норм русского языка / Под ред. А. А. Леонтьева. М., 1977.

Тарасов Е. Ф. Тенденции развития психолингвистики. М., 1987.

Фрумкина Р. М. Вероятность элементов текста и речевое поведение. М., 1989.

Шабес В. Я. Событие и текст. М., 1989.

Щерба Л. В. О трояком аспекте языковых явлений и об эксперименте в языкознании // Щерба Л. В. Языковая система и речевая деятельность. Л., 1974.

Якобсон Р. Избранные работы. М., 1985.

ЛИНГВИСТИЧЕСКАЯ ПРАГМАТИКА И ЕЕ ПРИКЛАДНЫЕ АСПЕКТЫ

Лингвистическая прагматика изучает условия использования языка коммуникантами в актах речевого общения. Конкретно эти условия включают в себя коммуникативные цели собеседников, время и место речевого акта, уровень знаний коммуникантов, их социальные статусы, психологические и биологические особенности, правила и конвенции речевого поведения, принятые в том или ином обществе, и т. д. В процессе речевого общения коммуниканты пользуются тем или иным кодом (языком, по Ф. де Соссюру), а также, при необходимости, и другими знаковыми системами, в том числе паралингвистическими. Условия использования языка нередко подразделяют на контекст (лингвистические условия), конситуацию (экстралингвистические условия) и коэмпирию (уровень лингвистических и энциклопедических знаний коммуникантов). Однако все эти условия нередко называются единым термином «контекст», в который, таким образом, вкладывается самое разное содержание. Естественно, что прагматика должна учитывать то влияние, которое оказывает такой контекст на используемые в процессе общения языковые средства.

В лингвистической прагматике в настоящее время сложилось несколько фундаментальных направлений. Одно из них — **учение о речевых актах.** Оно связано с именем Дж. Остина и восходит к идеям Л. Витгенштейна (1985) о языковых играх. Суть этого учения сводится к мысли о том, что языковые высказывания существуют не только для того, чтобы сообщать о положении дел в мире и квалифицироваться как истинные или ложные, но также и для того, чтобы давать возможность людям осуществлять с их помощью те или иные действия. Такие высказывания Дж. Остин назвал «перформативными» и предложил их классификацию. Он ввел также важные понятия локутивного, иллокутивного и перлокутивного актов и дал их

толкования (Остин, 1986). Это направление исследований развивалось в дальнейшем в работах Дж. Серля, Д. Вандервекена (Searle, Vanderveken, 1985), Дж. Версурена (Verschueren, 1980), Д. Вундерлиха (Wunderlich, 1976), Дж. Лича (Leech, 1963), Т. ван Дейка (Dijk, 1981), К. Баха и Р. Харниша (Bach, Harnish, 1980), Г. Г. Почепцова (1981), Ю. Д. Апресяна (1986), Н. Д. Арутюновой, Е. В. Падучевой (1985), И. П. Сусова (1985) и др. Многие из перечисленных здесь авторов предложили собственные оригинальные классификации речевых актов. Однако никто из них не ставил перед собой задачу выявить все языковые средства, выражающие речевые акты. Попытку решения этой задачи впервые предприняли Т. Баллмер и В. Бреннен-штуль (Ballmer, Brennenstuhl, 1980). Они построили сложную, многоступенчатую классификацию для 4800 глаголов английского языка и опубликовали словарь. Вслед за ним был издан также семантический словарь английских глаголов речевых актов А. Вежбицкой, в котором все глаголы были разбиты на 37 исходных групп (Werzbicka, 1987).

В рамках рассматриваемого направления была осуществлена формализация теории речевых актов и заложены основы иллокутивной логики. Эта задача была выполнена Дж. Серлем и Д. Вандервекеном (Searle, Vanderveken, 1980).

Теория речевых актов получила довольно широкое распространение в прикладных сферах лингвистики. Основное внимание здесь было сосредоточено на изучении речеактовой организации текстов научных, научно-технических, рекламных и деловых документов.

Второе направление исследований относится к **правилам и конвенциям** речевого общения. Речевые акты при этом выступают как своего рода алфавит, к которым и применяются данные правила.

Речевое общение между коммуникантами всегда ведется для достижения какой-то цели. Такая цель может заключаться в том, чтобы получить или сообщить сведения, договориться о проведении какого-нибудь мероприятия, обменяться мнениями о состоянии какого-нибудь дела или просто приятно провести время. Но если имеется цель, то должны существовать и средства ее достижения. Таким средством является вербальный обмен — диалог в устной форме или переписка в случае использования письменной формы общения. Вербальный обмен должен быть определенным образом организован. Эта организация регулируется разными правилами. Среди них наиболее важными являются четыре группы правил, а именно:

1) правила организации речевого взаимодействия;
2) правила организации дискурса;
3) правила организации информационного обмена;
4) правила учета статусных ролей коммуникантов в общении.

Правила организации речевого взаимодействия наиболее четко проявляют себя в диалоге. В формальном отношении диалогическое общение представляет собой координированную последовательность реплик двух сторон, обеспечиваемую принципом смены ролей, обусловленной тем, что наличие какой-то общей задачи, которую надо решить коммуникантам, не позволяет одной из сторон монополизировать свое право на речевую деятельность. Такое право необходимо предоставить и партнеру по диалогу или нескольким партнерам в случае полилога. В противном случае общая задача может остаться нерешенной. Однако доля вербального участия каждой из сторон в общении может быть различной и зависит от информационного дисбаланса сторон. Наличие информационного перевеса у одной из сторон обычно влечет за собой бо́льшую долю вербального участия. Однако ни доля вербального участия, ни общая задача не должны переоцениваться. Например, в случае пресс-конференции сторона, организующая ее, сообщает далеко не всю информацию, которой она располагает и которую ожидают репортеры. Информируемая сторона в таких случаях обычно испытывает информационный дефицит, сплошь и рядом являющийся источником разнообразных слухов. Формально пресс-конференция представляет собой серию вопросно-ответных диалогов.

При диалогическом общении с информационным перевесом одной из сторон возможна и такая ситуация, когда другая сторона максимально самоустраняется от активного участия в диалоге, создавая лишь видимость смены ролей с помощью сигналов типа «да-да», «разумеется», «еще бы», «надо же», «вы только подумайте», «вот тебе раз» и т. д. Такие вставки выполняют чисто стимулирующую функцию и не являются сигналами смены ролей в диалоге. Однако некоторые средства, например, перебивы, взгляды, жесты, замечания и т. д., могут выступать и как сигналы смены ролей. Особенно большую роль в этом отношении играет взгляд.

При спонтанном диалогическом общении стороны обычно мало контролируют или вообще не контролируют грамматической отмеченности высказываний, поэтому спектр отклонений от грамматической правильности в таких случаях довольно велик.

Правила организации дискурса в наибольшей степени проявляются в монологической речи и особенно в тексте, однако они приложимы и к диалогическому общению. От текста требуют прежде всего цельности и связности, т. е. такой организации высказываний (речевых актов и пропозиций), при которой текст отражал бы или конструировал некоторый мир в форме упорядоченной последовательности картин. Такая организация обеспечивается большим числом правил, объединяющихся вокруг тех или иных принципов.

Общий принцип — это линейная организация дискурса, обу-

словливающая перспективно-ретроспективную, или векторную, упорядоченность его содержания. В соответствии с этим принципом начальные фрагменты текста ориентированы перспективно, а конечные — ретроспективно. Однако у этого общего правила имеется масса исключений, которые позволяют комбинировать содержание не только в соответствии с этой общей схемой. Кроме того, существуют более частные правила структурирования текста, в частности, основанные на тех принципах, которые описаны Дж. Личем. К их числу относятся: 1) принцип перерабатываемости, который упорядочивает текст в соответствии с целями его более простого декодирования со стороны читателя. Этот принцип регулирует тема-рематическую структуру высказываний текста; 2) принцип ясности, исключающий неоднозначность и не допускающий линейного разрыва семантически связанных компонентов; 3) принцип экономии, обеспечивающий оптимальный объем текста без утраты ясности; 4) принцип выразительности, обеспечивающий реализацию эстетических особенностей текста (Leech, 1983).

Отдельные принципы могут противоречить друг другу. В этом случае перевес получает тот принцип, соблюдение которого наиболее уместно в данном контексте. Например, соблюдение принципа выразительности может нарушить принцип экономии, так как для повышения выразительности некоторого фрагмента текста может потребоваться синтаксически неоправданное, но эстетически необходимое повторение всего фрагмента или его части, в то время как принцип экономии запрещает всякие плеоназмы. Чрезмерная же экономия может нарушить ясность, а стремление любой ценой соблюдать ясность может прийти в противоречие с принципом перерабатываемости.

Разумеется, названными принципами правила организации дискурса не исчерпываются. Существуют и другие правила, которые неоднократно описывались в литературе. К ним относятся, например, правила формальной и содержательной связности отдельных предложений в рамках абзаца или сверхфразового единства, получившие названия правил когезии и когеренции. Сюда входят и правила объединения абзацев в связный текст.

Правила организации информационного обмена находят свое выражение в принципе кооперации, сформулированном Г. Грайсом (1985), а также в принципах вежливости, интереса и др., которые были описаны Дж. Личем (Leech, 1983).

Принцип кооперации был сформулирован преимущественно к условиям устного диалогического общения, однако с определенными поправками он может быть распространен и на текстовую коммуникацию, в рамках которой наиболее близка к диалогу переписка. Этот принцип регулирует количество и качество информации, выдаваемой каждым из коммуникантов на том или ином шаге речевого общения. Например, правило, ре-

гулирующее количество, требует, чтобы говорящий не скрывал нужной информации, но и не сообщал той, которая не требуется. Разумеется, этот объем всегда связан с фазой коммуникации. Правило, относящееся к качественному аспекту, запрещает выдачу ложной или непроверенной информации. Наконец, имеются правила, регулирующие уместность и общую организацию информации, включая требования ясности, краткости, однозначности и упорядоченности. Нетрудно заметить, что эти правила перекликаются с четырьмя рассмотренными выше принципами организации дискурса. В этом нет ничего удивительного, потому что формы организации обмена не могут не быть связаны с формами организации дискурса.

Отдельные правила, формирующие принцип кооперации, могут нарушаться, однако эти нарушения не подрывают общего принципа кооперации. Это значит, что всякое отклонение от правила получает содержательную интерпретацию как некий имплицитный смысл, исправляющий видимость такого нарушения. Если бы этот скрытый смысл не предполагался, то следовало бы подвергнуть сомнению состоятельность самого принципа кооперации.

К правилам организации информационного обмена относятся также принцип интереса и принцип Поллианны. Принцип интереса означает, что вербальное общение развивается более динамично, если информация представляет интерес для коммуникантов. Роль интереса в вербальном общении высоко оценивается не только в исследованиях по лингвистической прагматике, например, в работах Дж. Лича (Leech, 1983), но и в прикладной лингвистике, например, Р. Шенком, М. Лебовицем и Л. Бирнбаумом (1983), разработавшими интегральную понимающую систему. Принцип Поллианны требует, чтобы содержание речевого общения удовлетворяло критериям оптимистического настроения коммуникантов. Психологи и психолингвисты в последнее время уделяют большое внимание изучению этого принципа.

Правила учета статусных ролей коммуникантов проявляются в прагматической категории вежливости и связаны с иерархическими и неиерархическими отношениями между коммуникантами. В применении к японскому языку А. А. Холодович выделял следующие виды иерархических отношений: 1) иерархию возраста, 2) иерархию пола, 3) иерархию служебного положения, 4) иерархию сферы услуг.

Эти иерархические отношения означают, что в вербальном общении один коммуникант может иметь более высокий статус, чем другой. Например, старший по возрасту выше, чем младший; мужчина выше женщины; начальник выше исполнителя; клиент выше представителя обслуживающего персонала. В отличие от Востока в Европе женщина имеет более высокий ста-

тус, чем мужчина, а в России клиент ниже члена обслуживающего персонала.

К неиерархическим, или симметричным, отношениям А. А. Холодович относит: 1) отношение родства (неродственник выше родственника), 2) отношение приятельства (незнакомый выше знакомого) (*Холодович*, 1979).

Общий принцип вежливости означает, что низший всегда должен быть более вежлив к высшему, в то время как высший по отношению к низшему может обходиться нейтральными формами, а в ряде случаев может позволить себе даже грубые формы.

Вежливость реализуется в отдельных правилах, к числу которых относятся правила такта, великодушия, одобрения, скромности, согласия и симпатии. При этом такт, скромность и согласие в большей степени требуются от низшего по отношению к высшему, а великодушие, одобрение и симпатия — от высшего по отношению к низшему. При общении между равными соблюдение правил может зависеть от других факторов. Например, взрослые коммуниканты, не знакомые друг с другом, обычно соблюдают правила такта и согласия. Разумеется, в области соблюдения правил весьма широки этнические колебания. Так, в литературе неоднократно отмечалось, что в условиях Японии большое значение имеет правило скромности как характерная черта японской вежливости. В целом вежливость рассматривается как средство поддержания социальной гармонии.

С принципом вежливости тесно связан принцип иронии, суть которого заключается в том, что он позволяет говорящему нарушать социальную гармонию, делая вид, что ее соблюдает, и позволяя адресату этот вид принимать за действительность. На прямое оскорбление может последовать деструктивная реакция со стороны оскорбляемого. Оскорбление же, закамуфлированное под безобидное ироническое замечание, дает возможность оскорбляемому сделать вид, что он не понял оскорбления, или ответить иронией на иронию, что в конечном итоге во многих случаях может предотвратить конфликт, т. е. гармонизировать отношения.

Рассмотренные выше принципы и правила формируют условия контекста для вербальной коммуникации и, следовательно, относятся к прагматике. Они показывают, как коммуниканты пользуются речевыми актами и пропозициями в процессе общения.

Третье направление исследований в лингвистической прагматике связано с характером **знаний и информационных потребностей** коммуникантов. Эти знания и потребности выступают как конкретное наполнение речевых актов и пропозиций, упорядочиваемых с помощью правил рассмотренного типа в форму коммуникативного процесса. Чтобы этот процесс начал-

ся, необходима та или иная побудительная сила. Такой побудительной силой является информационный дискомфорт. В это понятие здесь вкладывается достаточно широкое содержание, поскольку оно интерпретируется не только в смысле потребности в информации, но и в смысле любого повода к общению.

Знания принято делить на лингвистические и энциклопедические. Применительно к рассмотренному материалу можно сказать, что данные о речевых актах и о правилах организации речевого взаимодействия, дискурса и информационного обмена можно рассматривать как знания лингвистические, хотя и относящиеся к сфере прагматики. Что же касается пропозиций и правил, связанных со статусными ролями, то здесь дело обстоит сложнее. К лингвистическим знаниям можно отнести правила языкового выражения пропозиционального содержания, а также правила языкового выражения статусных ролей с помощью дейктических или иных средств. Конкретная же информация из той или иной предметной области, заполняющая пропозициональные схемы, является по своей природе энциклопедической. Именно эта информация и является объектом главного интереса со стороны пользователей при вербальном общении в рамках конкретной предметной области, т. е. при профессиональном общении. Вполне естественно, что наибольшего развития в настоящее время достигли искусственные интеллектуальные системы, связанные, в первую очередь, с энциклопедическими знаниями. К этим системам относятся информационно-поисковые системы фактографического типа, экспертные системы и автоматические обучающие системы.

Фактографическая система только выдает энциклопедические знания в ответ на информационный запрос; экспертная система не только выдает знания, но и способна консультировать специалиста, направляя его по оптимальному пути при решении той или иной задачи, например, задачи медицинской диагностики; обучающая система выдает или корректирует знания пользователя в ходе обучения. Обычный режим экспертных и обучающих систем — диалоговый. Фактографические системы обычно работают не в диалоговом режиме, но, в принципе, возможны диалоговые фактографические системы. Важно при этом отметить, что в системах рассмотренного типа лингвистическая информация также может быть представлена как энциклопедическая. Типичный пример — система обучения иностранному языку. Сходная ситуация наблюдается в информационно-поисковых системах: документельная информация может быть выдана фактографической системой в форме фактографической информации. Это свидетельствует о том, что грань между лингвистическими и энциклопедическими знаниями не является резкой.

В настоящее время к интеллектуальным системам предъ-

является требование прагматической ориентированности, т. е. *подстройки под информационные потребности* индивидуального пользователя. Информационно-поисковые системы этому требованию не удовлетворяют, так как *в них действует принцип* релевантности, т. е. ориентации на запрос. В наибольшей степени этому требованию соответствуют экспертные системы.

Однако одной прагматической ориентированности недостаточно, особенно при работе в диалоговом режиме. Общение с интеллектуальной системой в таком режиме должно приближаться к человеческому, т. е. быть достаточно естественным и комфортным. Рассмотренная выше типология речевых актов и правил общения позволяет расширить коммуникативный потенциал интеллектуальных систем не только в смысле диапазона их энциклопедических знаний, но и в смысле их прагматической *подстройки под пользователя. Эти системы* должны быть не только способны осуществлять любой речевой акт, но они должны быть диалогически *гибкими, кооперативными*, вежливыми, т. е. должны достаточно точно моделировать речевое поведение человека.

Литература

Апресян Ю. Д. Перформативы в грамматике и словаре // Изв. АН СССР. Сер. лит. и яз. 1986. № 3, т. 45.

Витгенштейн Л. Философские исследования // Новое в зарубежной лингвистике. Вып. XVI: Лингв. прагматика. М., 1985.

Грайс Г. П. Логика и речевое общение // Новое в зарубежной лингвистике. Вып. XVI: Лингв. прагматика. М., 1985.

Остин Дж. Слово как действие // Новое в зарубежной лингвистике. *Вып. XVII: Теория речевых актов. М., 1986.*

Падучева Е. В. Высказывание и его соотнесенность с действительностью. М., 1985.

Почепцов Г. Г. Прагматика предложения // Иванова И. П., Бурлакова В. В., Почепцов Г. Г. Теоретическая грамматика современного английского языка. М., 1981.

Сусов И. П. Языковое общение и лингвистика // Прагматические и *семантические аспекты синтаксиса. Калинин, 1985.*

Холодович А. А. Грамматические категории уважительности в современном японском языке // Японское языкознание. М., 1979.

Шенк Р., Лебовиц М., Бирнбаум Л. Интегральная понимающая система // Новое в зарубежной лингвистике. Вып. XII: Прикл. лингвистика. М., 1983.

Bach K., Harnish R. Linguistic Communication and Speech Acts. *Cambridge, 1980.*

Ballmer Th., Brennenstuhl W. Speech Act Classification. Berlin e. a., 1981.

Dijk T. van. Studies in the Pragmatics of Discourse. The Hague e. a., 1981.

Leech G. N. Principles of Pragmatics. London e. a., *1983.*

Searle J. R., Vanderveken D. Foundations of Illocutionary Logic. Cambridge e. *a., 1985.*

Verschueren J. On Speech Act Verbs. Amsterdam, 1980.

Wierzbicka A. English Speech Act Verbs: A Semantic Dictionary. New York, 1987.

Wunderlich D. Studien zur Sprechakttheorie. Frankfurt-a/M., 1976.

ПРИКЛАДНОЕ ТЕРМИНОВЕДЕНИЕ И ЕГО НАПРАВЛЕНИЯ

Особое внимание при изучении прикладной лингвистики должно быть уделено прикладному терминоведению. Терминоведение представляет собой современную научную дисциплину, предметом которой являются термины и терминологические системы (терминосистемы). Оно возникло в середине XX в. на стыке лингвистики, логики, информатики, психологии и ряда других наук. Поскольку термины в любом языке — это в основном слова или словосочетания, которые подчиняются практически всем закономерностям создания и функционирования слов и их сочетаний, предмет и методы терминоведения в значительной мере совпадают с предметом и методами лингвистики. В частности, как и лингвистика, терминоведение включает в себя два раздела — теоретический и прикладной.

Поэтому, говоря о прикладном терминоведении, нужно прежде всего уточнить статус этого раздела терминоведения. Как в лингвистике имеется теория прикладной лингвистики, так существует и теория прикладного терминоведения и целый ряд направлений практической терминологической деятельности. При этом теоретические предпосылки практической терминологической деятельности являются различными: выбор направлений последней определяется той или иной терминоведческой школой, в равной мере как и применяемые здесь принципы, методы, объем, уровни деятельности, ее результаты и внедрение этих результатов в практику, например, в народное хозяйство, в экономику. Так, отечественная терминоведческая школа развивает следующие направления практической деятельности: лексикографическая терминологическая деятельность; унификация терминов и терминосистем, включающая в себя три вида — упорядочение, стандартизацию и гармонизацию терминов, работа редактора над терминами (терминологическое редактирование); перевод терминов в рамках перевода научных, технических и иных специальных текстов; создание терминоло-

гических банков данных (ТБД); организационно-методическая (научно-организационная) деятельность терминологических органов и центров на разных уровнях — от отраслевого до международного.

Перечень направлений практической терминологической деятельности не является ни окончательным, ни исчерпывающим. Дело в том, что сфера использования результатов терминологической деятельности ежедневно ставит перед терминоведением новые задачи, вызывающие к жизни новые направления работы. Так, необходимость и реальная возможность автоматизации хранения и обработки огромных объемов специальной информации недавно привели к созданию ТБД. Кроме того, в этой деятельности главную роль играют сами потребители терминологии (терминологической информации). В этом плане можно напомнить о таком интересном, но малоизученном направлении терминологической деятельности, как стихийно возникающие дискуссии разработчиков терминологических стандартов и рекомендаций об идентичном понимании терминов, а также работа создателей новой техники и технологии и научных сотрудников по присвоению названий вновь открытым или реализуемым специальным объектам и явлениям и по определению соответствующих понятий. Можно было бы утверждать, что присвоение названий (номинация) — это еще одно, причем важнейшее, направление практической терминологической деятельности, включающее в себя создание новых терминов с применением законов семантики, словообразования, синтаксиса, перевода (калькирования, транслитерации). Однако такой подход представляется не вполне оправданным. Во-первых, процессы номинации имеют место не только в сфере терминов (разница лишь в объекте номинации), и закономерности присвоения названий в принципе едины для всех сфер определенного естественного языка (так что это не специфически терминологическая деятельность). Во-вторых, в присвоении названий переплетаются терминологические, лингвистические и логические подходы, и дискуссии часто идут, по существу, о понятиях, а не о терминах, так что это опять-таки не терминологическая деятельность. В-третьих, и это, пожалуй, самое главное, нередко в этих процессах сознательно или бессознательно применяются знания закономерностей терминообразования, и поэтому лучше говорить о том, что терминологическая номинация — это теоретическая деятельность, имеющая практический выход (при том, что здесь есть сочетание управляемых и неуправляемых процессов), состоящий в создании терминов и введении их в речь и в терминосистемы.

Практическая терминологическая деятельность имеет два взаимосвязанных аспекта: оптимизационный и нормативный. С одной стороны, вся она направлена на обоснование выбора оп-

тимальных, т. е. наиболее соответствующих своей цели терминов и определений их значений, которые удовлетворяют потребность адекватного отражения объектов, процессов и признаков в специальных сферах. С другой стороны, отобранные оптимальные термины и определения закрепляются нормативными документами и предписаниями (стандарты, терминологические рекомендации, нормативные словари, редакторские указания и др.). Эти документы и предписания имеют бо́льшую силу, чем нормы, используемые в нетерминологической сфере лексики, и здесь еще одно отличие терминоведения от лингвистики, поскольку терминоведение имеет дело с нормативными документами, которых не знает лингвистика, и эти документы обладают правовым статусом, а термины в них используются наравне с производственными нормами.

Таковы некоторые общие особенности прикладного терминоведения и направлений деятельности в его рамках. Прежде чем коснуться отдельных направлений этой деятельности, нужно подчеркнуть, что все они взаимосвязаны, переходят друг в друга (от словаря к стандарту и т. п.), определяют друг друга.

1. Первое направление — это научно-техническая лексикография. Оно рассмотрено в следующем разделе данной книги.

2. Итогом объединения оптимизационной и нормативной функций практической терминологической деятельности явилось одно из основных направлений прикладного терминоведения — унификация терминов и терминосистем. Продуктами унификации терминов в настоящее время служат регламентирующие словари, ненормативные и нормативные: стандарты на термины разных категорий и сборники рекомендуемых терминов, которые отличаются друг от друга степенью обязательности, но сближаются по методам разработки и достижения унифицированности включенных в них терминов.

К началу 90-х годов XX в. в мире действовало более 20 тысяч стандартов на термины и определения. Основная их часть — это национальные стандарты, которые разрабатываются во всех промышленно развитых странах. Кроме того, существуют международные стандарты (их около двухсот), региональные стандарты, стандарты фирм и международных организаций. Большинство требований терминологических стандартов является обязательным для применения в различных видах документов, есть и рекомендуемые требования.

Упорядочение терминов реализуется в форме сборников рекомендуемых терминов. Они в основном разрабатываются в нашей стране Комитетом научно-технической терминологии Академии наук. К 1992 г. их было более 110.

Вопреки точке зрения, отрицающей целесообразность унификации терминов и закрепления их специальными документами, унификация терминов необходима для научно-технического прогресса, как необходима унификация узлов, деталей, типо-

размеров машин и механизмов. Ограничения во времени, в пространстве и в функциях, налагаемые на нормативные документы по унифицированной терминологии, устраняют опасность «остановки» научной деятельности, которая якобы произойдет, если будет устранена синонимия терминов. Известно, что при изложении результатов этой творческой деятельности упорядоченные или стандартизованные термины не применяются или применяются в ограниченном объеме.

Не может быть оспорена еще одна положительная функция унификации терминов — систематизирующая. В стандартах и сборниках рекомендуемых терминов эта функция проявляется еще сильнее, чем в словарях, в том числе систематизирующих. Разработка стандартов и сборников рекомендуемых терминов, в результате чего каждое место в терминосистеме занимается, как правило, одной лексической единицей и каждой лексической единице ставится в соответствие одно определение (правая часть), конечно, ведет к систематизации всех знаковых средств, относящихся к определенной сфере знаний или деятельности, к разграничению этих сфер и установлению связей между ними (через общие термины, используемые в левой и правой частях словарных статей). Однако — и это важнейшее положение в обосновании работы по унификации терминов — необходимо, чтобы эта работа всегда ставилась на прочную теоретическую базу таких наук и научных дисциплин, как теория стандартизации, лингвистика, логика, науковедение.

В этой связи следует сказать, что в последние годы начинает все шире применяться третий вид унификации терминов и терминосистем — их гармонизация на национальном и международном уровнях. Не стремясь к достижению совершенно одинаковых терминов в разных языках, да и в разных терминосистемах, относящихся к близким областям знания, специалисты-терминоведы и представители предметных наук ограничиваются взаимной увязкой терминов и терминосистем в разной степени.

При этом они учитывают различие самих языков, в которых фигурируют термины, — способы словообразования, фонетику, связи слов и т. п., а также обращают внимание на то, что каждая терминосистема зависит, в частности, от определенной теории, концепции, системы взглядов. К примеру, в физике существуют по крайней мере две терминосистемы — одна, в основе которой лежит классическая теория механики Ньютона, и вторая, основанная на теории относительности Эйнштейна; поэтому один и тот же термин, например *масса*, в этих терминосистемах имеет разное значение.

Отсюда следует, что при гармонизации терминосистем должны приниматься во внимание две группы факторов: лингвистические, т. е. связанные с особенностями тех языков, к которым относятся термины, и экстралингвистические, т. е. связанные с предметной областью и с теорией, которая ее описывает.

Экстралингвистическими предпосылками международной гармонизации терминосистем являются: интеграция знания, интернационализация науки и технологии, единство теоретической и методологической базы науки и технологии, характерные для современного этапа мировой цивилизации. В качестве лингвистических факторов, способствующих международной гармонизации терминосистем, могут быть названы следующие: формирование языков для специальных целей, достаточно близких друг другу по содержанию и форме представления в нескольких естественных языках, и активные процессы интернационализации терминов, выступающих в функции лексических единиц языков для специальных целей.

Существуют два основных принципа международной гармонизации терминосистем: а) она возможна и целесообразна, если в двух или более странах развита одна и та же область знаний или деятельности; б) она возможна, если в этих странах в основе данной области лежит одинаковая (идентичная) или близкая теория с одинаковой или близкой системой понятий.

В случае реализации в той или иной мере этих принципов и с учетом языкового воплощения их в терминах могут быть достигнуты три степени гармонизации терминосистем:

1) общность систем понятий при частичном совпадении теорий и различии языковых средств (в этом случае гармонизация выступает лишь в форме взаимной увязки систем понятий и обозначающих их терминов, например, в теории надежности в нашей стране и во Франции);

2) общность систем понятий при полном совпадении теории и различии языковых средств (в этом случае гармонизация имеет вид единства систем понятий и структурно-семантической близости обозначающих их терминов, например, в области авиационной терминологии в нашей стране, Великобритании, Франции и Германии);

3) общность систем понятий при полном совпадении теории и языковых средств (в этом случае гармонизация реализуется в идентичных по содержанию, объему и характеру связей системах понятий и интернационализации терминов, например, в некоторых разделах химии, генетики, медицины на большинстве европейских языков).

Прикладными результатами международной гармонизации терминосистем являются: международные терминологические словари и стандарты, многоязычные информационно-поисковые тезаурусы. В частности, деятельность по международной гармонизации терминологических стандартов занимает важное место в работе Международной организации по стандартизации (ИСО) и служит предметом проекта ИСО 860 «Международная гармонизация понятий и терминов».

3. Перевод научных, технических и иных специальных

терминов является еще одним направлением практической терминологической деятельности. Можно сказать, что при переводе специального текста самым важным служит правильный, адекватный перевод входящих в него терминов. Не случайно на многочисленных конференциях и семинарах переводчиков проблемы выбора способов перевода терминов занимают основное место. К настоящему времени разработаны рекомендации по применению этих способов (*Лейчик, Шелов*, 1989—1990). Вкратце они могут быть сформулированы следующим образом.

а) Оптимальным способом перевода служит выявление в языке перевода (ЯП) эквивалента термина языка оригинала (ЯО). Применение этого способа возможно в тех случаях, когда страны, в которых распространены ЯО и ЯП, достигли одного и того же уровня общественного развития или прошли этот уровень в какой-то период своей истории. Так, социально-экономический немецкий термин Bauer переводится термином *крестьянин*, а русский термин *рабочий* переводится английским термином worker. Французский физический термин vapeur переводится русским термином *пар*, а русский термин *управление* — немецким термином Leitung. Особенностью этих терминов является то, что обычно термины-эквиваленты разных языков уже существовали в обоих языках, и в момент перевода их выявляют, находят, а не создают, как это бывает с новыми терминами, неологизмами.

б) Новый термин в ЯП может быть создан путем придания существующему в этом языке слову или словосочетанию нового значения под воздействием термина ЯО. Русское слово *дерево* (ботанический термин) получило новое значение и стало термином теории классификации, теории управления и информатики, в частности в терминах-словосочетаниях *дерево целей* и *поисковое дерево* под влиянием соответствующего англ. tree. Русское слово *род (мужской род, средний род)* получило значение лингвистического термина под влиянием латинского термина genus (точно так же и в теории классификации: *род* — genus, *вид* — species и т. д.). Ярким примером может служить также русский политический термин *разрядка*; это слово, применявшееся вначале как физический термин, обязано своим новым политическим значением французскому détente от глагола détendre — «*ослаблять, разряжать*».

в) Если структура переводимой лексической единицы в обоих языках совпадает, мы имеем дело с так называемой семантической калькой. На основе вышеприведенного английского термина tree (*дерево*) создан термин tree structure, который передается на русском языке термином *древовидная структура* (а не *древесная структура* или *структура дерева*, как можно было бы ожидать). При использовании семантического калькирования структура термина, создаваемого в ЯО, соответствует нормам ЯО, а структура термина, создаваемого в ЯП, соответ-

ствует нормам ЯП. Общей является только семантика терминов обоих языков, почему этот способ перевода и называется семантическим калькированием. В технических науках поэлементный перевод (калькирование) сложных по структуре терминов также широко распространен: англ. swithing diagram — *коммутационная схема,* motor selector — *моторный искатель* (из государственных стандартов на термины электроники).

г) Если структура лексической единицы заимствуется при переводе вместе с этой единицей, мы имеем дело со структурной калькой, или собственно калькой. Греческий термин philosophia еще в средние века был передан в русском языке термином *любомудрие,* термин, состоящий из греко-латинских элементов, televisio переводится на немецкий язык термином das Fernsehen. Это поэлементный перевод сложной лексической единицы, при котором каждому элементу из ЯО соответствует элемент в ЯП, сопровождается появлением в ЯП новой, чуждой ему модели. К примеру, в русском языке наряду с моделью сложного существительного «глагольная основа + именная основа», по которой построено слово *любомудрие,* более была распространена модель «именная основа + глагольная основа», откуда и слово *мудролюбие* (ср. *свободолюбие, блюдолиз, книгочей*). Некоторые из заимствованных моделей (элементов структуры) закрепляются в результате калькирования в ЯП, другие остаются чужеродными, существуют только в изолированных структурных кальках. В этом и преимущество, и недостаток калькирования.

д) Когда в процессе перевода заимствуются и семантика, и структура, и форма (звуковой состав и написание) термина, мы имеем дело с заимствованием. При этом следует четко различать заимствования, зависящие от непосредственных контактов двух языков, и интернационализмы, в первую очередь, построенные из греко-латинских элементов и определяемые традиционными особенностями европейской культуры, основанной на классическом образовании.

Что касается собственно заимствований, то отношение к ним при переводе неоднозначно. В тех случаях, когда термин «приходит» в ЯП вместе с новым понятием, которое он выражает, или с новым объектом, который он обозначает в ЯО, то заимствование термина может быть допущено. Так, в русский язык из американского варианта английского языка перешел термин *джинсы* (анг. jean — комбинезон; джинсовая ткань; jeans — джинсы), из русского языка было заимствовано во французский язык слово *самовар* и др. Заимствование этой так называемой безэквивалентной лексики при переводе логически и методически оправданно. В равной мере это относится и к сфере духовной жизни: русский термин *патриот* из франц. patriote, англ. soviet из русского *совет.* Однако в тех случаях, когда уровень науки и техники в странах, в которых применяются

ЯО и ЯП, одинаков, переводчик и редактор должны попытаться либо найти эквивалент термина ЯО в ЯП, либо построить новый термин из элементов ЯП. Например, вряд ли была необходимость заимствовать английский термин know-how — *ноу-хау*, если его можно перевести термином *секреты производства*.

е) В ряде случаев термин должен переводиться описательной конструкцией. Этот способ перевода применяется, прежде всего, для безэквивалентных терминов, отражающих реалии определенной страны. Так, из целого ряда вариантов перевода советского термина *ударник* на английский язык самым правильным оказалась описательная конструкция, помещенная в словаре Уэбстера для студентов: a worker who had excelled in voluntary increase of production — «рабочий, который на добровольных началах добился высоких показателей в производстве».

4. *Терминологическое редактирование* представляет собой один из видов научного редактирования текстов наряду с литературным и техническим редактированием. Рассматривая терминологическое редактирование, нужно охарактеризовать его цель, задачи, принципы и методы.

Цель его, как и при других видах редактирования, — добиться от текста результатов, при которых можно было бы получить максимальный эффект в познавательном, образовательном или воспитательном смысле.

В ходе терминологического редактирования эта цель достигается путем работы над специфическим материалом — терминами, которые, как было сказано, зависят и от определенного языка, и от места в той или иной терминосистеме. Поэтому задача терминологического редактирования — это достижение правильного использования терминов и слов, их заменяющих, с точки зрения их роли в конкретном тексте и с точки зрения места термина среди других терминов данной отрасли науки или техники, внутри данной терминосистемы.

Принципы терминологического редактирования в значительной степени определяются стилем и жанром редактируемого текста. Например, в сугубо научном тексте — статье, монографии — очень много специальных терминов, и здесь не дается их объяснение: считается, что читатели хорошо знают эти термины. А в научно-популярной статье или в газетном очерке термины могут заменяться описательными оборотами, так как читатели, может быть, впервые знакомятся с научной проблемой и политическими или экономическими событиями.

Что касается методов терминологического редактирования, то чаще всего применяются два метода. Первый заключается в том, что редактор приводит имеющиеся в тексте термины в соответствие с терминами и определениями, либо утвержденными нормативными документами (стандарты, сборники рекомендуемых терминов), либо широко распространенными в дан-

ной области науки и техники. Второй метод состоит в том, что если автор редактируемого текста настаивает на вводимых им новых терминах, то он должен дать в этом же тексте объяснение терминов, их толкование, определение.

5. Одно из новейших направлений терминологической деятельности — создание ТБД. Оно выросло из работ по формированию терминологических картотек и связано с быстрым ростом количества научных, технических, экономических и других терминов и невозможностью вручную обрабатывать огромные объемы терминологической информации. Компьютеризация, являющаяся в настоящее время важнейшим способом обработки значительных объемов различной специальной информации, снижает трудоемкость и повышает качество информационных работ, сокращает сроки обработки и выдачи информации. Особое место в этой области занимают работы по обеспечению специалистов и ученых точной и закрепленной нормативными документами терминологией. Именно потребность в сборе, хранении и выдаче такой терминологии и привела к созданию в нашей стране и за рубежом многочисленных ТБД — автоматизированных хранилищ терминов; при этом каждый термин в ТБД снабжен дополнительной информацией: указывается, в частности, каково значение термина, т. е. приводится определение, даются эквиваленты термина на других языках, краткие формы термина, его синонимы, допустимые и недопустимые, сообщается, каким документом утвержден термин (например, национальным или международным стандартом), и т. д.

В зависимости от цели создания ТБД их можно разделить на две группы: ориентированные на обеспечение работ по переводу научно-технической литературы и документации и предназначенные для обеспечения информацией о стандартизованной и рекомендованной терминологии.

В настоящее время существует несколько десятков крупных ТБД. У нас в стране имеется банк стандартизованных терминов в органах стандартизации (в институте ВНИИКИ), банк медицинских терминов и некоторые другие. Одним из самых больших в мире является ТБД германской фирмы «Сименс», который содержит около 2,5 млн. терминологических записей на 8 языках, в том числе на русском. Созданы ТБД во французской организации по стандартизации, два банка в Канаде. В Германии ТБД функционируют в Институте стандартизации и в других органах управления. В Люксембурге в терминологическом бюро Комиссии Европейского союза создан крупнейший международный ТБД в виде автоматизированного терминологического словаря на нескольких языках; он носит название Евродикаутом.

За ТБД большое будущее, из автоматизированных хранилищ они превращаются в средство проведения серьезных на-

учных исследований, не говоря уже о том, что они имеют большое практическое значение (Лингвистическая концепция..., 1989).

6. Наконец, к прикладному терминоведению можно отнести организационно-методическую деятельность. Она состоит в руководстве созданием терминологических стандартов и словарей, в формировании ТБД, в проведении конференций, семинаров и других встреч по проблемам терминологии, в обучении основам терминоведения различных категорий специалистов — переводчиков, стандартизаторов, научных работников в сфере прикладной лингвистики.

Все эти и другие формы организационно-методической терминологической деятельности выполняют национальные и международные органы, часть которых специально предназначена для работы с терминологией, а другая часть делает это наряду с выполнением иных функций.

Собственно терминологической деятельностью занимаются следующие органы. В Москве расположены: Научно-исследовательский институт комплексной информации по стандартизации и качеству (ВНИИКИ), который отвечает за проблемы стандартизации терминологии, а также выпускает единственный в нашей стране терминологический журнал «Научно-техническая терминология»; Комитет научно-технической терминологии (КНТТ) РАН, который руководит работами по упорядочению терминологии в области естественных и технических наук; издательство «Русский язык» и другие издательства, которые выпускают специальные словари; научно-производственное объединение «Медицинская энциклопедия», которое создает банк медицинских терминов и выпускает медицинские терминологические словари и справочники. Работают в Москве и общественные организации, в том числе постоянно действующий семинар по методологическим проблемам научно-технической терминологии, семинар по проблемам медицинской терминологии и др. Большую работу по созданию ТБД и терминологических словарей выполняют Санкт-Петербургский университет и Омский терминологический центр. Нижегородский университет периодически выпускает сборники научных трудов по проблеме «Термин и слово».

Велики достижения терминоведов в независимых государствах: в Украине существует Комитет научной терминологии при Академии наук, крупные специалисты имеются в Черновцах, Ужгороде, Днепропетровске (в университетах этих городов). Комитет научной терминологии Академии наук Латвии постоянно работает над упорядочением латышской терминологии и выпускает сборники терминов на латышском, немецком, английском и русском языках.

Из международных организаций следует назвать Международный центр информации по терминологии (Инфотерм), ко-

торый расположен в Вене, Международную организацию по унификации терминологических неологизмов (МОУТН), которая находится в Варшаве и издает журнал «Neoterm». Разработка международных терминологических стандартов входит в обязанности Международной организации по стандартизации (ИСО), Международной электротехнической комиссии (МЭК) и др.

Все сказанное позволяет сделать вывод о том, что в настоящее время во всем мире уделяется большое внимание терминологической деятельности, как в теоретическом, так и в практическом плане. Это связано с тем, что ученые и производственники осознали тот факт, что от рациональной, точной, отработанной терминологии в значительной степени зависит успех их научной, технической, коммерческой и экономической деятельности.

Литература

Герд А. С. Основы научно-технической лексикографии: Как работать над терминологическим словарем. Л., 1986. 73 с.

Гринев С. В. Введение в терминологическую лексикографию: Учеб. пособие. М., 1986. 103 с.

Даниленко В. П., Волкова И. Н., Морозова Л. А., Новиков Н. Р. Лингвистический аспект стандартизации терминологии. М., 1993.

Квитко И. С., Лейчик В. М., Кабанцев Г. Г. Терминоведческие проблемы редактирования. Львов, 1986. 151 с.

Краткое методическое пособие по разработке и упорядочению научно-технической терминологии. М., 1979. 126 с.

Лейчик В. М., Шелов С. Д. Лингвистические проблемы терминологии и научно-технический перевод: В 2 ч. Ч. 1. М., 1989. 40 с; ч. 2. 1990. 78 с.

Лингвистическая концепция Терминологического банка данных Машинного фонда русского языка: (Проект) / Под ред. проф. А. С. Герда. М., 1989. 104 с.

Разработка стандартов на термины и определения: Рекомендации. Р. 50-603-1-89. М., 1990. 54 с.

НАУЧНО-ТЕХНИЧЕСКАЯ ЛЕКСИКОГРАФИЯ

Лексикография — это теория и практика разработки и создания словарей разных типов. Нередко именно лексикографию наряду с орфографией, орфоэпией, основами культуры речи, с методикой обучения языку называют среди основных практически важных аспектов языкознания.

Объектом научно-технической лексикографии является теория и практика создания специальных (терминологических) словарей. В специальных словарях толкуется и определяется терминология и номенклатура отдельных наук и отраслей знания, представленная в книгах, журналах, а также в деловых документах министерств, ведомств, вузов, в производственной документации заводов, фабрик, комбинатов, сельхозобъединений.

Виды специальных словарей. Специальные словари делятся на словари, адресованные человеку, и словари, ориентированные на ЭВМ. К первым относятся одноязычные и переводные терминологические словари, словари идентификаторов, частотные словари терминов, рубрикаторы, классификаторы; ко вторым — информационно-поисковые тезаурусы для информационных систем разных типов, а также разные типы собственно машинных словарей (основ, оборотов, аффиксов), применяемых в различных системах автоматической переработки текста *.

В отличие от обычных толковых словарей любого языка терминологические словари предназначены прежде всего для специалистов в данной области знания или отрасли техники.

По назначению выделяются, с одной стороны, терминологические словари, ориентированные на специалистов с высшим образованием, профессиональных переводчиков научной литературы. Такие словари должны давать возможность специалистам достаточно высокой квалификации найти в словаре как само слово, так и то понятие, которое им обозначается, различные синонимы, встречающиеся в той или иной статье или кни-

* В принципе любой словарь может быть введен в ЭВМ.

ге; не только сделать сам перевод статьи с другого языка, но при необходимости также уточнить и расширить данные словаря через обращение к его источникам, приведенным в словарной статье. Именно поэтому здесь желательны большие одноязычные и многоязычные толково-переводные терминологические словари, построенные по идеографическому принципу. С другой стороны, выделяются одноязычные и переводные словари, адресованные студентам, специалистам среднего звена в промышленности и сельском хозяйстве, работникам радио и печати. Их главная цель — обеспечить только перевод специального текста. Это относительно простые по структуре нормативные алфавитные терминологические словари среднего и малого объема. Особо выделяются терминологические словари для школьников. Здесь главным фактором является возраст, на который ориентирован словарь, и он определяет способ подачи материала, обязательное наличие элементов увлекательности, рисунков, схем, конкретных примеров. В последнее время появились научно-популярные терминологические словари, а также значительное число сугубо научных словарей, которые отражают терминосистемы отдельных областей науки. Как правило, эти словари строятся не по алфавитному, а по систематическому принципу.

Также по назначению выделяются словари узуса (регистрирующие, инвентаризирующие), регламентирующие, учебные, информирующие, систематизирующие.

К регистрирующим словарям, словарям узуса, относится большинство существующих терминологических словарей. Они представляют состояние терминологии в данной отрасли знания таким, как оно сформировалось, сложилось к настоящему времени.

Регламентирующие словари делятся на ненормативные и нормативные. Наиболее яркими примерами нормативных словарей являются стандарты на термины и словари (сборники) рекомендуемых терминов.

Стандарт на термины и определения — это терминологический словарь системного типа, имеющий официальный, государственно-правовой статус. Терминологический стандарт состоит из следующих частей: вводной, основной, алфавитных указателей. В вводной части указывается область распространения стандарта, уточняется объект стандартизации, а также приводятся сведения о недопустимых терминах, синонимах, эквивалентных, иллюстрациях и приложениях.

В основной части стандарта находим термины и определения, которые располагаются в соответствии с принятой для данного стандарта понятийной системой. Материал здесь представлен в двух графах: первая графа имеет заголовок «Термин», а вторая — «Определение». В первой графе приводится номер термина, стандартизованный термин, краткая форма (ес-

ли термин является словосочетанием), недопустимые к употреблению синонимы стандартного термина (с пометой «ндп»), эквиваленты стандартизованного термина на английском, немецком и французском языках. Определения понятий помещаются в графе «Определение» и могут сопровождаться также примечаниями или иллюстративными материалами (графиками, рисунками, чертежами) (Волкова, 1984).

Приведем фрагмент из ГОСТа «Коррозия металлов».

Термин	Определение
Общие термины	
1. Коррозия металлов...	Разрушение металлов вследствие химического или электрохимического взаимодействия их с коррозионной средой...
2. Коррозионная стойкость	Способность металла сопротивляться коррозионному воздействию среды

К терминологическим словарям относятся и сборники рекомендуемых терминов по различным отраслям знаний, регулярно выпускаемые Комитетом научно-технической терминологии. Приведем фрагмент из сборника рекомендуемых терминов по основным понятиям магнитной термодинамики.

13. Диффузия магнитного поля	Проникновение магнитного поля внутрь сплошной среды, обусловленное ее конечной проводимостью.

В источниках этого типа содержатся не только термины, рекомендуемые к употреблению в качестве основных, но также и их определения, синонимы и переводные эквиваленты.

Разработка нормативных терминологических словарей крайне актуальна. Нормативный терминологический словарь — словарь, содержащий термины, которые относятся к одной терминосистеме и обязательный характер применения которых подтвержден специальным распорядительным документом организации, для нужд которой разработан словарь. Содержанием нормативного словаря может являться терминосистема четко очерченной сферы знаний или деятельности, причем сформулированная с точки зрения той или иной теории (в этом плане политехнический или другой энциклопедический словарь не может быть нормативным). Степень обязательности применения терминов нормативного словаря может быть различной: его единицы могут быть рекомендуемыми, как в сборниках рекомендуемых терминов, стандартизованными, как в стандартах на термины разных уровней (от фирменных до государственных, международных), гармонизованными, как в международных многоязычных стандартах и словарях типа «Международного электротехнического словаря».

Особо по назначению выделяются словари информирующие и систематизирующие. Примером информирующих словарей является информационно-поисковый тезаурус.

Информационно-поисковый тезаурус — одноязычный контролируемый и изменяющийся словарь лексических единиц, основанный на лексике одного естественного языка, отображающий семантические отношения между лексическими единицами и предназначенный для обработки и поиска информации.

Информационно-поисковые тезаурусы — это словари, облегчающие поиск информации при ее автоматической обработке. ИПТ максимально раскрывают смысловые, семантические отношения между терминами.

Основной единицей ИПТ являются термины-дескрипторы. Алфавитная лексико-семантическая часть ИПТ представляет собой свод дескрипторных словарных статей. Каждая дескрипторная словарная статья начинается термином-дескриптором, при котором ниже в пределах статьи приводятся синонимы этого дескриптора, а также другие термины, связанные с основным дескриптором родо-видовыми или ассоциативными отношениями. Приведем пример из ИПТ по сталям:

Сталь, выплавленная в дуговых печах
с Электродуговая сталь
в Электросталь
н Сталь вакуумно-дугового переплава
Сталь плазменно-дугового переплава
Сталь, выплавляемая в плазменно-дуговых печах
см. Сталь плазменно-дугового переплава

Здесь с — синоним, в — вышестоящий термин, н — нижестоящий термин.

К систематизирующим специальным словарям относятся классификаторы, рубрикаторы, частотные словари, словари терминосистем.

В классификаторах в строго упорядоченном (систематически и алфавитно) виде приводятся сотни названий объектов науки, техники и их характеристики. Здесь следует назвать прежде всего классификаторы технико-экономических показателей, классификатор промышленной и сельскохозяйственной продукции, а также такие классификаторы, как классификатор-ценник сырья и материалов, классификаторы предприятий и организаций, инструментов и оснастки.

По строению и по типу описываемых данных к классификаторам близки отраслевые каталоги изделий промышленности, номенклатурные справочники, каталоги промышленных образцов, сборники оптовых цен на товары и изделия, сборники экспресс-информации, выпускаемые различными министерствами и ведомствами.

Приведем фрагмент из классификатора промышленной и сельскохозяйственной продукции.

Условные коды				Молотки отбойные
31	4196	1000	10	Молотки отбойные типа МО
31	4196	1006	04	Молоток отбойный МО-6П ГОСТ 22044-76
31	4196	1007	03	— МО-7П То же
31	4196	1008	02	— МО-5П То же
31	4196	2000	06	Пневматические молоты
31	4196	2010	04	Пневматический молот типа ПЛ
31	4196	2011	03	Пневматический молот типа ТУ 24-8-421-76 ПЛ 1М...

Существенное место среди таких терминологических словарей занимают рубрикаторы. Рубрикатор отрасли — опубликованный или машинописный документ, являющийся вариантом классификации объектов, понятий, относящихся к данной отрасли и состоящий из рубрикаций (перечней рубрик), определяющих тематику разделов рубрикатора. Это методический документ, предназначенный для использования при решении конкретных задач, возникающих в системе информационного обслуживания специалистов отрасли и предприятий. Рубрикатор служит руководством для описания информационных потребностей специалистов и предприятий, определения политики комплектования справочно-информационных фондов в отраслевой системе научно-технической информации, координации обработки информации предприятия отрасли, определения тематических границ информационных массивов автоматизированной информационно-поисковой системы, формулирования информационных запросов.

Чаще всего рубрикаторы содержат систематизированные уже перечни основных направлений отрасли, важнейшие типы объектов и процессов. Приведем фрагмент из «Рубрикатора по приборостроению».

Порядковый индекс рубрики	Наименование рубрики	Аспект рассмотрения объекта	Индекс УДК
905	Измерительная техника и оборудование научно-исследовательских лабораторий и служб контроля производственных процессов		
9051	Приборы для исследования свойств атомов и молекул	Б, Д — И, Н	542, 2/539, 18—539, 19—539, 69
9052	Электронные микроскопы	Б, Д — И, Н	621, 385, 833

Отраслевые частотные словари интересны тем, что в них представлены частоты отдельных терминов.

Одним из самых существенных признаков словаря является тематический охват терминов. Если словарем охвачено несколько предметных областей, то мы имеем дело с многоотраслевым

словарем. Часто, однако, создаются отраслевые и узкоотраслевые словари; их разработка вызвана современной дифференциацией наук.

Для словаря важным признаком является наличие правой и левой частей словарной статьи: слева — заголовочное слово, справа — то, что о нем говорится в словаре. По содержанию левой части словарной статьи различаются собственно словари терминов и словари терминоэлементов. К последним относятся, например, словари интернациональных морфем. Среди словарей, выделяемых по содержанию правой части словарной статьи, необходимо отметить обычные переводные словари, словари, в которых даются определения соответствующих терминам понятий (они традиционно называются толковыми).

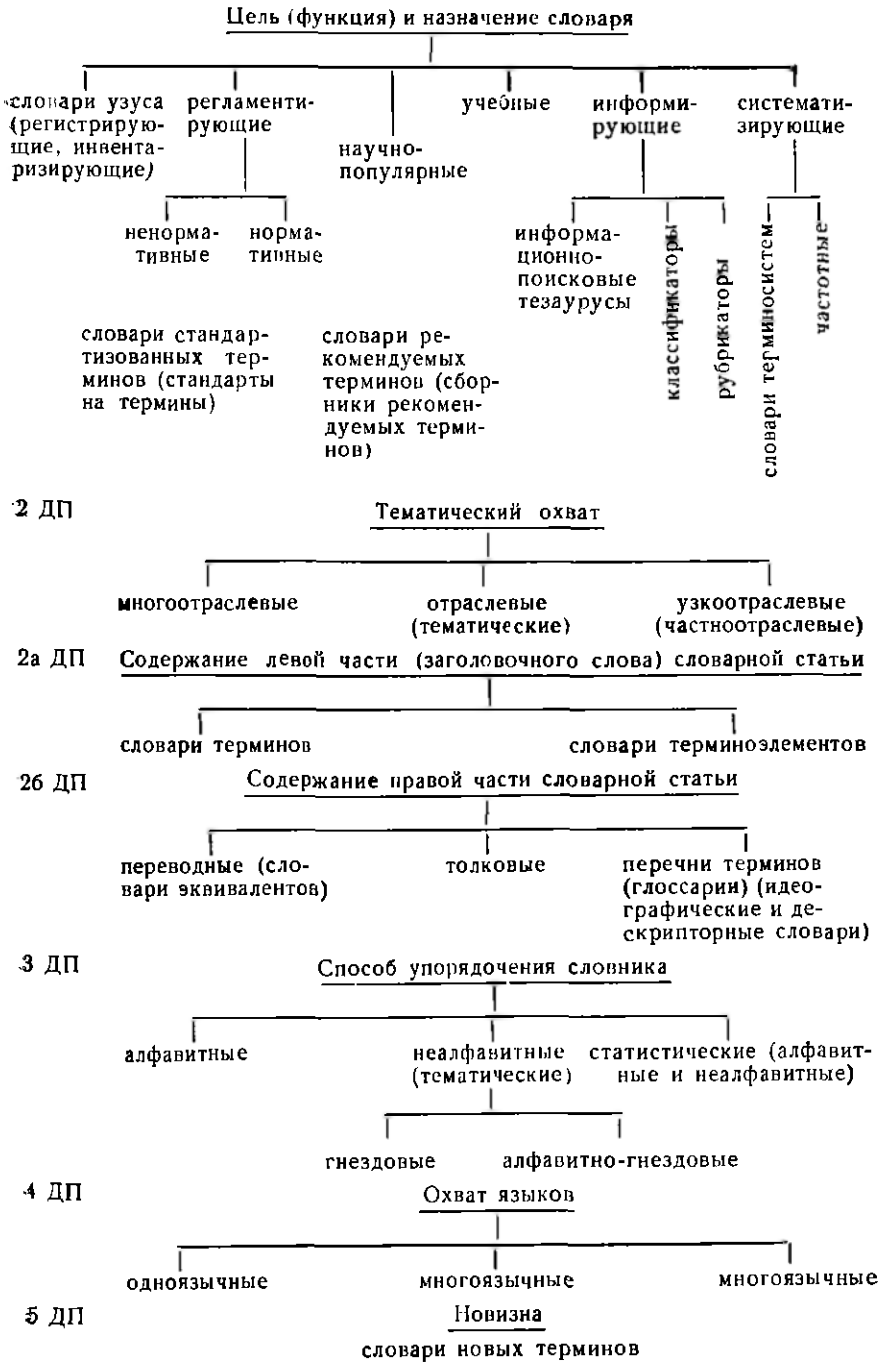
Классы толковых терминологических словарей являются наиболее важными в практике создания отраслевых и узкоотраслевых словарей. Основные задачи разработки толкового словаря следующие. Толковый терминологический словарь должен служить справочным пособием в конкретной области, отражать систему ее понятий и содержать все термины, необходимые для научной и практической работы специалистов. Он должен в терминах отразить все новые понятия, возникающие в связи с последними достижениями науки и техники, и дать определение этих понятий, позволяющее специалистам правильно использовать термины, обозначающие новые понятия. В идеографических и дескрипторных словарях (информационно-поисковых тезаурусах) правая часть может быть прямо не представлена; здесь в самом подборе терминов заложен принцип их отбора и группирования — например, по классам условной эквивалентности в тезаурусах.

По способу упорядочения словника в терминологических словарях кроме традиционного алфавитного и гнездового применяется статистический способ (по признаку частоты встречаемости в текстах).

По охвату языков терминологические словари делятся на одноязычные и переводные (двух-, трехязычные и многоязычные). Особым основанием деления терминологических словарей является новизна термина; в этом плане нужно упомянуть многочисленные «Тетради новых терминов», в которых приводятся английские, французские, немецкие термины новейших областей знания и их русские эквиваленты.

В целом классификация специальных словарей на основе дифференциальных признаков (по В. М. Лейчику) представлена на схеме (см. с. 293).

Источники специальных словарей. Основными источниками терминологических словарей разных типов являются обобщающие монографии и статьи, относящиеся к данной отрасли знания, принадлежащие перу ведущих ученых и специалистов-практиков. В числе привлекаемых источников важное


Цель (функция) и назначение словаря
словари узуса (регистрирующие, инвентаризирующие)
регламентирующие
научно-популярные
учебные
информирующие
систематизирующие
ненормативные
нормативные
информационно-поисковые тезаурусы
классификаторы
рубрикаторы
словари терминосистем
частотные
словари стандартизованных терминов (стандарты на термины)
словари рекомендуемых терминов (сборники рекомендуемых терминов)
2 ДП
Тематический охват
многоотраслевые
отраслевые (тематические)
узкоотраслевые (частноотраслевые)
2а ДП
Содержание левой части (заголовочного слова) словарной статьи
словари терминов
словари терминоэлементов
2б ДП
Содержание правой части словарной статьи
переводные (словари эквивалентов)
толковые
перечни терминов (глоссарии) (идеографические и дескрипторные словари)
3 ДП
Способ упорядочения словника
алфавитные
неалфавитные (тематические)
статистические (алфавитные и неалфавитные)
гнездовые
алфавитно-гнездовые
4 ДП
Охват языков
одноязычные
многоязычные
многоязычные
5 ДП
Новизна
словари новых терминов

место занимают также крупные вузовские учебники, так как именно в них обычно содержатся наиболее ясные и четкие определения понятий и терминов науки. В терминологические словари должны быть включены и все термины данной отрасли, представленные в энциклопедиях и других словарях.

Существенным источником лексики терминологических словарей являются и такие вторичные документы, как рефераты и аннотации книг и статей, опубликованные в реферативных журналах, а также описания изобретений и патенты, деловые документы. В сжатом и концентрированном виде они дают относительно полное представление об основных терминах, функционирующих в литературе отрасли. Приведем пример описания изобретения.

(51) МКИ Н 05 11/00

Г. Г. Канаев, Э. Г. Фурман (71). Научно-исследовательский институт ядерной физики, электроники и автоматики при Томском политехническом институте имени С. М. Кирова.

(54) (57) Линейный индукционный ускоритель, содержащий охватываемые витками ферромагнитные сердечники с уложенной поверх них двойной полосковой линией, состоящей из четырех электродов, многоканальный разрядник, источник напряжения, вакуумный тракт с фокусирующей катушкой, отличающийся тем, что с целью уменьшения веса и габаритов и повышения КПД ускорителя два потенциальных электрода двойной полосковой линии, подключенные к источнику напряжения и многоканальному разряднику, уложены в спираль по образующей цилиндра, а два других выполнены в виде цилиндров, подключенных к виткам ферромагнитных сердечников, концы которых разнесены на 180°.

Нетрудно заметить насыщенность терминами, которая характерна для таких документов. Например, только из последнего документа составители терминологического словаря могут извлечь такие термины, как *линейный индукционный ускоритель, ферромагнитный сердечник, полосковая линия, электрод, многоканальный разрядник, источник напряжения, вакуумный тракт, фокусирующая катушка, КПД, спираль, цилиндр.*

Нельзя не учитывать и такой источник терминологических словарей, как «Тетради новых терминов», регулярно издаваемые Центром переводов научно-технической литературы и документации.

Немало терминов и номенклатурных обозначений содержат также многочисленные отраслевые и ведомственные руководства и инструкции по эксплуатации и ремонту оборудования, технические условия, технические рекомендации, технические описания, рекомендации по технике обслуживания, паспорта машин, приборов, руководства по сборке, указатели технической документации, типовые проекты, памятки рабочему по использованию оборудования.

В ходе работы над словарем из различных источников выбираются термины и формируется словник терминологического словаря. Словник терминологического словаря — это реестр единиц (слов, словосочетаний, сокращений, символов, сочетаний

слов и символов), выбранных из источников словаря, предназначенных для включения в словарь и подлежащих определению и описанию. Обычно словник терминологического словаря строится чисто интуитивно, путем отбора в словарь так называемых основных, важнейших терминов с точки зрения отдельных специалистов. Такой подход, естественно, создает большой простор для субъективизма, императивизма и, как правило, не опирается на анализ научных текстов.

Хронологические рамки специальных словарей. Но прежде чем приступить к непосредственной работе по формированию словника, автор будущего терминологического словаря должен определить хронологические рамки подготавливаемого издания.

Для новых отраслей знания, таких как вычислительная техника, радиофизика, биофизика, этот вопрос не столь актуален, так как любой терминологический словарь по указанным отраслям обычно представляет их лексику со времени не столь давнего появления этих наук. Иное положение в таких естественных науках, как физика, механика, математика, химия, биология, география, геология, а также и в большинстве гуманитарных. Русские тексты этих наук насчитывают по 150—200 лет своего исторического развития, и здесь вопрос о хронологическом периоде, отраженном в терминологическом словаре, оказывается одним из наиболее серьезных.

В широком историко-филологическом и историко-научном плане обсуждаемый вопрос — это вопрос правильного прочтения, понимания и перевода с одного языка на другой научных текстов предшествующих периодов. Небезразличен он и для перевода современной научно-технической литературы. В ряде текстов и в наши дни нередко встречаются старые термины.

Специалистам в области фундаментальных наук часто бывают нужны терминологические сведения из работ предшественников, из исследований разных школ, времен и направлений. Современные авторы и сегодня могут примыкать к тем или иным концепциям, которые высказывались ранее, развивать их далее в своем творчестве. Во многих работах мы находим обращение к истории вопроса, критику и пересмотр различных идей. При этом стремление к научной точности диктует, как правило, педантичное, дословное употребление терминов авторов-предшественников. Встречаются книги, статьи, полностью посвященные полемике с предшественниками. Обращение к именам, понятиям и концепциям XIX — начала XX в. нередко в филологии, математике, биологии, химии; осуществляется оно через язык науки. Именно поэтому в основе отбора лексики для терминологических словарей должна быть периодизация истории языка науки в тесной связи с историей самой науки.

Так, например, история русских научных биологических текстов делится на следующие периоды: 1) XVIII в. — 50-е годы XIX в. — период начального становления русских биологиче-

ских текстов, 2) 50—60-е годы XIX — начало XX в. (10—20-е годы), 3) с 10—20-х годов XX в. до наших дней. Следовательно, при работе над терминологическим словарем по биологии терминологию биологических работ, созданных до начала XX в., не следует относить к современному языку биологии.

По-видимому, в каждом конкретном случае могут быть разные решения. Конечно, тому, кто пользуется терминологическим словарем, удобнее всего иметь под рукой один словарь, в котором были бы собраны все возможные сведения о данном термине, включая и его историю. Наличие такого полного словаря исключило бы необходимоть обращения к разным словарям. Однако есть два серьезных соображения против таких терминологических словарей, которые охватывали бы глубокие хронологические пласты в истории «своей» терминологии и одновременно служили бы справочниками по современной терминологии. Во-первых, это достаточно сложный и постоянно меняющийся характер научных знаний, сложная эволюция понятийной иерархии, постоянные изменения содержания и объема понятий, перераспределение терминов между понятиями, неустойчивость синонимических рядов в терминологии. Во-вторых, необходимо дать читателю прежде всего максимум сведений по современной терминологии и при этом обычно с ориентацией на лексические параллели в других языках. Все это говорит о том, что создание терминологического словаря, содержащего сведения по истории развития «своей» терминологии, представляет собой задачу отдельную и особую.

Однако в любом случае, даже при создании терминологического словаря, ориентированного на современную терминологию, остается открытым достаточно сложный вопрос о нижнем хронологическом пределе словаря. Будут ли это 40—50-е или 20—30-е годы XX в., или же рубеж XIX и XX вв. — решать можно только применительно к каждой конкретной отрасли знания, науке, исходя из основных этапов развития ее языка.

В идеале решение вопроса о хронологических рамках терминологических словарей должно опираться на периодизацию истории науки, на анализ истории языка данной науки в их взаимосвязи, что, в свою очередь, лишний раз демонстрирует общефилологический, историко-культурный аспект многих прикладных проблем.

Одним из путей, ведущих к уменьшению субъективизма при формировании словника терминологических словарей, является предварительное моделирование логико-понятийной системы, характерной для данной науки, отрасли знания. Логико-понятийный анализ специальных текстов — необходимое предварительное условие создания терминологических словарей.

Структура специальных словарей. Среди вопросов, связанных со структурой терминологических словарей, выделяются два основных: 1) о принципах организации данных в тер-

минологическом словаре, 2) о структуре отдельных фрагментов терминологического словаря.

О принципах организации данных в терминологических словарях. Так же как и в классической лексикографии, в научно-технической лексикографии выделяются следующие общие принципы организации словарного материала: алфавитный, алфавитно-гнездовой, идеографический. Абсолютное большинство терминологических словарей строится по алфавитному принципу. Этот принцип давно и заслуженно оправдал себя как наиболее удобный для пользования способ размещения слов в любом словаре.

В то же время многочисленные практические потребности специалистов свидетельствуют о том, что терминологический словарь может быть построен не только по алфавитному или алфавитно-гнездовому принципу расположения терминов, но и по идеографическому принципу, на логико-понятийной основе, на базе строгой научной классификации понятий. Этот последний принцип, в конечном счете, соответствует логике развития данной науки, отрасли знания и дает возможность легче найти термин по его значению. Так, специалистам в области отдельных наук часто бывает нужно знать, как, каким словом называется то или иное понятие, с каким понятием связаны те или иные синонимы. В исследовательской практике специалиста и в практике перевода может возникнуть ситуация отождествления данного термина с аналогичным по форме термином из работ давних лет.

Основой структуры терминологического словаря, построенного по идеографическому принципу, на строгой логико-понятийной основе, является не алфавитный список терминов в левой части словаря, а выраженное в вербальном, условном или формализованном виде понятие, занимающее свое, строго определенное место в общей логико-понятийной структуре словаря. Например, в тезаурусе Роже (Roget's thesaurus..., 1978) заголовком словарной статьи является код понятия и ключевое слово, выражающее это понятие, за которым после тире следуют указания на часть речи, синонимы:

380. Cold — N. coldness, etc. adj.; low temperature; frigidity, coolness...

В немецком идеографическом словаре Дорнзайфа (Dornsejf, 1958, S. 211) также каждая словарная статья начинается номером понятия в данной тематической группе, и далее следуют синонимы, близкие друг к другу в структурном отношении:

1. Zeiträume <...>
-lang: studenlang, jahrelang.

О структуре отдельных фрагментов терминологических словарей. Чаще всего словарная статья, согласно общей теории лексикографии, состоит из следующих частей: 1) заголовок сло-

варной статьи; 2) грамматическая характеристика термина; 3) характеристика термина по его употребительности, стилистической окраске и происхождению; 4) семантическая характеристика термина.

Структура словарной статьи терминологического словаря, построенного по алфавитному принципу, близка к структуре ее в толковых словарях и энциклопедиях. Заголовком словарной статьи здесь обычно является слово, стоящее на своем алфавитном месте. В качестве заголовочного слова в терминологическом словаре могут выступать не только термины-слова, но и словосочетания, комбинации слов и символов, аббревиатурные сокращения.

Терминологический словарь всегда рассматривается специалистами как важный шаг на пути нормализации и унификации терминологии, как основное средство систематизации научных и технических понятий отрасли, и поэтому любой терминологический словарь в идеале всегда словарь нормативный. И здесь главное — выделение и продуманная подача основных общеупотребительных, стандартизованных терминов, отделение и особое расположение синонимов к основному термину. Особо выделяются термины, которые могут быть рекомендованы в качестве стандартизованных. В ряде случаев, особенно в отраслевых и межотраслевых терминологических словарях — геологических, биологических, медицинских, военных и т. п., применяются пометы, уточняющие сферу употребления термина. Так, в «Толковом словаре английских геологических терминов» в квадратных скобках отмечается, когда термин преимущественно употребляется не столько в геологии, сколько в смежных науках; применяются такие пометы, как *астрон., биол., бот., геодез., гидрол., палеонт., гляциол.*, например:

axis [палеонт.], calving (гляциол.).

Указания на этимологию термина приводятся во многих терминологических словарях, например, в «Геологическом словаре»:

ТРАВЕРТИН (итал. travertino).
ТЮРИНГИЙ (по Тюрингской обл.).

Обычны в словарях указания на греческое или латинское происхождение терминов.

Семантическая характеристика — основная и самая сложная часть словарной статьи любого словаря. В терминологических словарях, построенных по алфавитному принципу, семантическая характеристика термина осуществляется либо через логическое определение его значения, либо через переводной эквивалент в переводных терминологических словарях. При формулировании определений в терминологическом словаре руководствуются следующими правилами.

1. Термин и его определение должны быть соразмерны, т. е. все понятия и объекты, называемые с помощью термина, должны быть охвачены определением, и наоборот, все понятия и объекты, которые подразумевают определение, должны включаться в понятия и объекты, называемые с помощью термина. Например, если определить *центрифугу* как аппарат для разделения жидких неоднородных систем, то такое определение будет неправильным (слишком широким), так как оно охватывает и фильтры, которые не относятся к центрифугам. Следовательно, здесь термин и его определение несоразмерны, ибо определение задает более широкий класс объектов, чем термин. Правильным в приведенном случае является определение центрифуги как аппарата для разделения жидких неоднородных систем под действием центробежных сил.

2. Признаки, используемые в определении термина, должны быть наиболее важными и существенными для данной тематической области. Так, если определять *металлы* как простые вещества, обладающие блеском, то такое определение будет не подходящим для металловедения, ибо видовой признак (признак обладания блеском) несуществен для данной тематической области. Уместным в данном случае будет признак, характеризующий специфику строения атомов металла. Это требование особенно актуально для определения терминов, используемых одновременно в разных дисциплинах и тематических областях. Так, термин *золото* может иметь различные определения (и, следовательно, разные родовые понятия и видовые признаки) в геологии, металловедении, химии, экономике.

3. В определении не должно быть порочного круга, т. е. термин не должен определяться с помощью другого термина, который, в свою очередь, определяется, в конечном счете, через первый.

4. Термины, используемые в определении, должны быть либо уже определенными в данном словаре, либо хорошо известными и однозначно понимаемыми. Если в определении используются термины, относящиеся к смежным дисциплинам или дисциплинам, более общим, чем та, по которой составляется словарь, то желательно в специальном приложении разъяснить понятия, которые эти термины выражают в данном определении, или указать источник, где эти термины и соответствующие понятия зафиксированы.

5. Для определения значения термина исключительную роль играет осознание места термина в системе научных понятий, в терминосистеме. Определения терминов должны быть системными, т. е. должны отражать место понятия, выражаемого одним термином, во всей системе понятий — наряду с понятиями, выражаемыми другими терминами. В частности, один и тот же признак, который лежит в основании деления понятий, выраженных разными терминами, должен выражаться в определени-

ях этих терминов однотипными языковыми конструкциями. Например, термины *газовый лазер, жидкостный лазер* и *твердотельный лазер* выражают различные соподчиненные понятия, в основании деления которых лежит признак — вид активного элемента лазера. Поэтому неправильной (несистемной) является совокупность определений: «газовый лазер — это лазер, в котором активной средой является газ или смесь газов», «жидкостный лазер — это лазер с жидким активным элементом», «твердотельный лазер — это лазер, в котором рабочим веществом является твердое тело». Правильной (системной) в данном случае будет, например, такая совокупность определений: «газовый лазер — это лазер с газовым активным элементом», «жидкостный лазер — это лазер с жидким активным элементом», «твердотельный лазер — это лазер с твердотельным активным элементом».

6. Определения должны быть выражены преимущественно словами; математические и химические формулы, а также иллюстрации чаще всего служат подсобным материалом.

7. Содержание определений, применяемых в различных языках для одного термина, может отражать определенное различие в толковании отдельных понятий в разных странах, что зависит от существования нескольких теорий, описывающих область знания. Это следует иметь в виду, когда переводится определение из иностранных толковых терминологических словарей. В данном случае необходимо специально оговорить источник определений.

Толковые терминологические словари дают точное логическое определение научного понятия, отнюдь не передавая всего его содержания. На практике обычно в каждом терминологическом словаре мы встречаем и определения, и дополнительные описания. Так, в терминологическом словаре «Горное дело» и в «Геологическом словаре» каждый термин сопровождается определением, за которым следует краткое описание понятия. Сочетание краткого и четкого логического определения с лаконичным описанием (толкованием) представляет собой идеальную структуру семантизации термина в терминологическом словаре.

Вскрытие семантики научных текстов невозможно без привлечения специалистов соответствующих отраслей знания. При этом описание понятий — дело специалиста, а лаконичное, компактное и более точное формулирование соответствующих терминологических значений — дело лингвистов вместе со специалистами.

К построению полного лексикографического определения термина могут быть выдвинуты следующие требования: предварительный анализ логико-понятийной схемы, выделение родового понятия, выделение существенных признаков понятия, выде-

ление отличительных признаков данного понятия от других понятий.

Определение значения терминов в терминологическом словаре желательно проводить по единой, заранее заданной системе параметров для каждого вида понятия. Так, например, в определении терминов одного типа могут быть заданы такие признаки, как назначение, устройство, состав, область применения, место установки и т. д. Примеры из терминологического словаря по промышленной трубопроводной арматуре.

Арматура трубопроводная. Устройства, предназначенные для управления потоками рабочих сред (жидкой, газообразной, газожидкостной, порошкообразной суспензии и т. п.) путем изменения проходного сечения. А. Т. устанавливается на трубопроводах и емкостях...

Задвижка, и, *ж.* Трубопроводная арматура, в которой запирающий или регулирующий орган перемещается возвратно-поступательно к оси потока рабочей среды. З. используют как чисто запорную арматуру, т. е. запирающий орган (клин) обычно находится в крайних положениях «открыто» или «закрыто»...

В переводном терминологическом словаре переводные эквиваленты также представляют собой семантическую характеристику термина входного языка, так как в таком словаре значение термина раскрывается обычно средствами родного для читателя языка. Верно подобранный эквивалент семантически должен полностью соответствовать тому научному понятию, которое передает переводимое слово. Тем более ответственно следует подходить к подбору переводных эквивалентов. Они не только должны быть отобраны из наиболее авторитетных первичных источников, но и должны полностью соответствовать терминоупотреблению, принятому среди специалистов. Вопрос этот далеко не безразличен для таких языков, как испанский, английский, где употребление тех или иных терминов для одних и тех же понятий может сильно различаться от страны к стране. Именно в связи со сказанным, например, в «Словаре названий морских промысловых рыб морской фауны» для большинства испанских, английских названий рыб по источникам проставлена страна, в которой употребляется данный термин.

Нередко переводчику специальной литературы важно не только найти точный эквивалент нужного слова в своем языке, его синонимы, но и знать план его содержания шире, а именно, действительно ли данное слово, приведенное в словаре, обозначает то же самое понятие, что и слово в тексте оригинала. Переводчику важно знать не только само слово, но и то, что этим словом обозначается, с каким именно понятием данное слово связано. Таким образом, в идеале и переводной терминологический словарь должен содержать не только слова входного языка и их переводные эквиваленты, но и определение того понятия, с которым данное слово соотносится, списки тех синонимов, которые соотносимы именно с данным понятием (а не с другим) в этом языке, в этой стране.

Так, например, рыбу пагр по-испански в Аргентине называют Besugo, а в Мавритании Lauriana; круглую сигарную ставриду на Кубе — Antonio, а в Венесуэле — Chupagaro. Рыбу луфарь в Аргентине и Венесуэле называют Anchoa, на Канарских островах — Anjova, Pez rez. Само по себе нахождение точного переводного эквивалента нередко может быть связано с привлечением некоторого минимума собственно фактографических данных. Таким образом, переводной терминологический словарь может быть не лишен и элементов энциклопедизма, заключающихся в кратких комментариях, наличии схем, рисунков. Схемы и рисунки следует использовать как можно шире в качестве важного дополнительного средства раскрытия значения термина. Так, они широко применяются в словарях по геологии, географии, технике.

Семантическая характеристика термина в одноязычном или переводном терминологическом словаре может быть дополнена и расширена путем отражения в терминологическом словаре синонимии, антонимии, родо-видовых и ассоциативных семантических отношений.

Синонимия отражается лишь в лучших, наиболее полных терминологических словарях. Так, в «Толковом словаре английских геологических терминов» нередко встречаем прямые указания на то, что этот термин — синоним такого-то другого, например:

endospore — эндоспор... Син. термина «интина» (intine)
felsite — фельзит... Син. aphanite; felstone
fenste — окно... Син. термина «тектоническое окно» (window)
inclusion — включение... Син. enclave, enclosure

Синонимы приводятся и в отечественных словарях «Горное дело» и «Геологический словарь», например:

Точка экстремальная — син. термина пиктермического эффекта.
Тип (серия, ряд) пород бореальный — син. термина породы (серия, ряд) арктического типа.

Гипонимия — это отношение включения, или отношение слова с более общим, или родовым, значением к слову с более частным, или видовым, значением. Гипонимию можно понимать и как отношение класса к подклассу. Гипонимическое поле имеет строго иерархический характер. В качестве родового должно быть выбрано слово с максимально общей семантикой, которому и подчиняются слова, обозначающие видовые понятия. Все эти слова должны быть расположены на одном уровне иерархии и должны разграничиваться по единому основанию. Этот ряд видовых слов образует своего рода субполе (субпарадигму) первой степени подчинения. То или иное слово в этом субполе может иметь собственные слова-подвиды. Так возникает иерархия. В случае многоуровневой иерархии каждое субполе следует рассматривать как отдельное поле, поскольку

каждый гипероним (родовое слово) является идентификатором по отношению к своим гипонимам (видовым словам). Все уровни иерархии, связанные с общей вершиной, выступают как макрополе по отношению к зависимым от гипонимов микрополям. Типичным примером может служить классификация семейств, родов и видов рыб, зверей, птиц и растений мира.

На одном уровне иерархии могут располагаться термины, семы, которые одновременно относятся как к одной семантической категории, так и к разным. В таком случае говорят о фасетной характеристике терминов. Чаще всего фасет (F) — это семантический признак, основание для классификации дескрипторов, занимающих один и тот же уровень иерархии. Пример из ИПТ по техническим средствам навигации:

Радиопеленгатор

F — место установки	F — метод индикации
Бортовой радиопеленгатор	Визуальный радиопеленгатор
Наземный радиопеленгатор	Слуховой радиопеленгатор
F — область применения	F — диапазон рабочих волн
Авиационный радиопеленгатор	ДВ-пеленгатор
Морской радиопеленгатор	КВ-пеленгатор

В этом примере разные типы фасетов (F) показывают различные аспекты семантической классификации терминов, подчиненных термину *радиопеленгатор*. При создании терминологического словаря методы фасетной характеристики терминов целесообразно использовать при разработке словарной статьи на основании опорных терминов типа *радиопеленгатор*. В такой словарной статье после определения термина можно привести и его семантические фасеты и основные синонимы, связанные с этими фасетами.

Отдельные фасеты могут объединяться в фасеты более общего порядка — субкатегории. Так, в результате анализа фасетов по сталям и их типам, по коррозии, а также фасетов по навигации выясняется, что все они объединяются в ряд субкатегорий — таких, как локализация, процессы, характеристики, параметры, назначение и т. д.

Для наиболее полного отражения гипонимических отношений в терминологическом словаре используются указатели иерархических семантических отношений между терминами. Приведем пример иерархического указателя из «Тезауруса научно-технических терминов»:

Приборы для определения ориентирных направлений

Астрономические приборы ориентирования
Гироскопические приборы ориентирования
Гирокомпасы
Гиротеодолиты
Магнитные приборы ориентирования
Буссоли (с. 409).

Пользуясь таким иерархическим указателем связей терми-

нологического словаря, легче определить и выявить термины, находящиеся в отношении гипонимии. Например, очевидно, что гирокомпасы — это разновидность гироскопических приборов ориентирования, а сами эти приборы — подвид приборов для измерения ориентирных направлений.

Терминологические словари по-разному представляют родовидовые отношения между терминами. Так, в «Геологическом словаре» видовые понятия чаще помещаются в круглые скобки при родовом, например:

Углеводороды метановые (алканы, парафины).

В этом же словаре иногда встречаем примечание «Частный случай. . .», например:

Точка инвариантная. . . Частный случай Т. и. — точка тройная.

Важным средством углубления семантической характеристики термина в терминологических словарях являются ассоциативные отношения. Это те отношения, которые не могут быть выражены обычной схемой иерархического подчинения. Так, например, в современном биологическом тексте немалое место занимают отношения, связанные с описанием морфологии вида, числа органов, их высоты, длины, ширины, расположения, с характеристикой воды по цвету. При описании числа органов, их длины, высоты, ширины наиболее важны такие отношения, как длина — возраст — вес, возраст — рост — пол, вес — длина — число особей, размер — возраст — темпы роста по годам. При описании частей тела встречаются не только все основные цвета, но и вся гамма их оттенков (желтоватый, темноватый, светло-оливковый, медно-красный, киноварно-красный). Ассоциативные термины, как и синонимы, отмечают лишь лучшие терминологические словари, преимущественно геологические. Так, в «Толковом словаре английских геологических терминов» все такие термины помещаются в конце словарной статьи, за пометами *ср.* или *см. также,* например:

humification — гумификация. . . см. также mor.
point bar — коса, вытянутая отмель. . . ср. channel bar, meander scroll, meander bar.

В «Геологическом словаре» ассоциативные термины обычно помещаются за пометой *см.*, например:

Тип рельефа. . . см.: Съемка геоморфологическая, картирование геоморфологическое.
Тип угля касьянитовый см. Касьянит.
Тюрингий. . . см. Цехштейн.

Иногда в одном и том же словаре такие термины приводятся не только за пометами *см.*, *см. также*, но и в самом определении, при раскрытии значения термина, например:

Tal, n — долина... В качестве собственно эрозионных долин (Erosionstal) следует рассматривать ущелье (см. Schlucht), каньон (см. Canyon), теснину (см. Klamm), а также промоины (Tobel) (Муравски, 1980).

Широко применяется этот путь показа терминов, ассоциативно связанных с описываемым термином, в словаре «Горное дело», например:

Кенорватель — приспособление для захвата и отрыва керна от забоя скважины и удержания его в колонковой трубе при подъеме бурового снаряда из скважины.

В ряде терминологических словарей приводятся термины-антонимы, что, естественно, расширяет круг семантически ассоциативных терминов, например, в «Толковом словаре английских геологических терминов»:

Permeability (геол.) — проницаемость. Син. perviousness. Прил. permeable. Ант. inpermeability.

В том или ином виде в терминологическом словаре должна быть отражена и представлена терминосистема данной отрасли знания, науки, техники.

Так, для того чтобы зафиксировать терминосистему в терминологическом словаре, необходимо: 1) провести подготовку проекта словника путем отбора кандидатов в термины; 2) построить логико-понятийную систему, в которой места понятий могут быть заняты кандидатами в термины, символами и другими типами знаков или быть свободными (являться лакунами); 3) построить терминосистему путем наложения проекта словника на систему понятий. При этом места понятий заполняются существующими терминами со всеми возможными их вариантами или специально созданными терминами; 4) произвести оценку терминов, входящих в терминосистему, и отбор вариантов, которые будут признаны нормативными (предпочтительными) и ненормативными, но включаемыми в словарь в качестве нерекомендуемых по разным причинам логического или лингвистического характера; 5) упорядочить словник словаря по алфавитному, систематическому или алфавитно-гнездовому принципу в зависимости от избранной цели и назначения словаря. При этом оптимальным является систематическое расположение терминов в словаре, что соответствует полииерархической структуре системы понятий.

В толковом терминологическом словаре терминосистема может отражаться через: 1) толкование, определение; определение значения должно учитывать показ необходимых связей и отношений между терминами данного терминологического словаря; 2) типизацию структуры определения, ввод типовых определений, обязательную представленность тех или иных семантических компонентов в определениях однотипных по характеру слов (указания на назначение, состав, область приме-

нения); 3) задание уровня и степени глубины отражения самой терминосистемы; 4) задание предела глубины семантизации отдельных терминов; 5) графическое выделение в тексте определения терминов, которые есть и толкуются в этом же терминологическом словаре; 6) разветвленную систему отсылочных определений; 7) указания на гипонимию; 8) сочетание определения и описания понятия; 9) указание на синонимы, антонимы; 10) указание ассоциативных семантических связей термина; 11) указание на лексико-семантическую группу; 12) справочный отдел словарной статьи; 13) пометы, уточняющие сферу употребления слова; пометы типа *устар., нов., ндп., неупотр.;* 14) историческую справку о некоторых толкованиях термина в прошлом; 15) схемы, графики, рисунки; 16) элементы гнездования, выделение опорных слов; 17) приведение всех производных при основном термине; 18) показ терминоэлементов и их типового употребления в терминах; 19) дополнительные входы в терминологический словарь в виде разного рода приложений (терминосистем, фасетов, тематических ключей).

Для терминологических словарей, построенных по алфавитному принципу, важным средством выявления семантических связей между терминами данной области знания, положения отдельного термина в терминосистеме может служить систематический указатель терминов, помещаемый в приложении к терминологическому словарю. В систематическом указателе все термины распределяются по тематическим классам (дескрипторным областям). Приведем фрагменты списка дескрипторов тематических классов из ИПТ по информатике.

КЛАСС № 5. ТЕКСТЫ. ПЕРЕВОД

Тексты. Речь

ЗАГЛАВИЯ
КОНТЕКСТ
НАЗВАНИЯ
ПРЕДЛОЖЕНИЯ
РЕЧЬ
СЛОВА

Словари

ПОМЕТЫ ЛЕКСИКОГРАФИЧЕСКИЕ
СЛОВАРИ
СЛОВАРИ АВТОМАТИЧЕСКИЕ
СЛОВАРИ МНОГОЯЗЫЧНЫЕ
СЛОВАРИ ОБРАТНЫЕ
СЛОВАРИ ЧАСТОТНЫЕ
СЛОВНИКИ

Разделы лингвистики

ЛЕКСИКОГРАФИЯ
ЛЕКСИКОЛОГИЯ
ЛИНГВИСТИКА МАТЕМАТИЧЕСКАЯ
ЛИНГВИСТИКА ПРИКЛАДНАЯ
ЛИНГВИСТИКА СТРУКТУРНАЯ
МОРФОЛОГИЯ

СЕМАНТИКА (НАУКА)
СИНТАКСИС...

Грамматика

ГРАММАТИКА
ГРАММАТИКА ЗАВИСИМОСТЕЙ
ГРАММАТИКА НЕПОСРЕДСТВЕННО СОСТАВЛЯЮЩИХ
ГРАММАТИКА ПОРОЖДАЮЩАЯ
ГРАММАТИКИ ТРАНСФОРМАЦИОННЫЕ
ГРАММАТИКИ ФОРМАЛЬНЫЕ...

Лексика и семантика

ЛЕКСИКА
МНОГОЗНАЧНОСТЬ
ОМОНИМИЯ...

Все термины, включенные ИПТ по информатике в класс 5, следует считать находящимися в ассоциативных отношениях между собой. При этом наибольшую степень семантической близости обнаруживают термины, входящие в подклассы (например, Тексты; Речь; Словари и т. д.).

Лексикография — область языкознания, которая наиболее тесно связана с практикой редакционно-издательского дела и полиграфией. Внутреннее построение словаря, размещение словарных статей, справочного аппарата, типы сокращений, графических знаков, шрифтов, выделений занимают немалое место в процессе создания и подготовки словаря. Именно поэтому любой словарь, в конечном счете, представляет собой итог совместной работы специалистов самых разных областей и профессий.

Литература

Алексеев П. М. Статистическая лексикография. М., 1975.
Андрющенко В. М. Машинный фонд русского языка. М., 1988.
Берков В. П. *Вопросы двуязычной лексикографии*. М., 1973.
Волкова И. Н. Стандартизация русской научно-технической терминологии. М., 1984.
Геологический словарь. Т. 1—2. М., 1978.
Герд А. С. Основы научно-технической лексикографии. Л., 1986.
Горное дело. М., 1981.
Гринев С. В. Введение в терминологическую лексикографию. М., 1986.
Комарова З. И. Семантическая структура специального слова и ее лексикографическое описание. Свердловск, 1991.
Муравски Г. Толковый словарь немецких геологических терминов. М., 1980.
Очередные задачи русской академической лексикографии / Под ред. Г. Н. Скляревской. СПб., 1995.
Тезаурус научно-технических терминов. М., 1972.
Теория и практика научно-технической лексикографии / Под ред. Ю. Ф. Майсурадзе. М., 1988.
Толковый словарь английских геологических терминов. Т. 1—3. М., 1977—1979.
Убин И. И. Автоматический переводной словарь. М., 1989.
Шайкевич А. Я. Проблемы терминологической лексикографии. М., 1983.

УЧЕБНАЯ ЛЕКСИКОГРАФИЯ

Учебная лексикография является одним из самых молодых разделов общей лексикографии, имеющей долгую и хорошо изученную историю. В самом деле, если работы по составлению словарей разных типов ведутся в учреждениях, существующих подчас не одну сотню лет, то первый в России центр учебной лексикографии — Сектор лексикографии и словарей Научно-методического центра русского языка при МГУ (ныне Сектор учебной лексикографии Института русского языка им. А. С. Пушкина) организационно оформился лишь в 1966 г. В настоящее время масштабы теоретической и практической работы по созданию русских одноязычных и русско-иноязычных словарей в нашей стране постоянно расширяются.

Специфической особенностью учебной лексикографии как научной дисциплины является ее тесная связь с общей лексикографией, лексикологией, семасиологией, психо- и социолингвистикой, а также с целым рядом проблем методики обучения языкам. Таким образом, кроме традиционных собственно лексикографических проблем учебная лексикография должна учитывать и многочисленные методические факторы: этап обучения, вид речевой деятельности, родной язык обучаемых, наличие или отсутствие языка-посредника, наличие или отсутствие языковой среды, возрастные особенности и уровень общей культуры обучаемых и т. д.

Только при одновременном и комплексном учете вышеперечисленных лингвистических и методических факторов учебная лексикография может обеспечить различные формы обучения русскому и иностранным языкам адекватными учебными словарями, дать правильные методические рекомендации по их использованию. Корректное применение принципов общей теории лексикографии позволяет устранить из преподавания лексики лингвистически устаревшие методы и приемы, указать границы

и возможности внутриязыковой и межъязыковой эквиваленции, выявить потенциал лексико-семантической сочетаемости, показать собственно лингвистические ограничения и области максимально эффективного применения каждого из этих приемов.

Учет методических факторов позволяет определить: 1) адресата словаря (школьники, студенты и т. д.); 2) тип словаря (одноязычный или переводной); 3) объем словника (словарь-минимум, средний или полный); 4) структуру словарной статьи, приемы и методы подачи лексического материала внутри словарной статьи (эквиваленция, дефиниция, иллюстрация, противопоставленность другим словам в лексико-семантической группе (синонимия, антонимия, паронимия, ассоциативность, сочетаемость, грамматические, орфоэпические и стилистические особенности и т. д.); 5) полиграфическое исполнение.

Правильный учет принципов общей теории лексикографии позволяет адекватно представить языковую реальность, а методические требования определяют самое существенное в учебном словаре — его нацеленность на изучение языка, применение на определенном этапе и для определенной формы обучения, т. е. его функцию.

В тесной взаимосвязи и взаимовлиянии лингвистики и методики заключается сущность учебной лексикографии как самостоятельной лингводидактической дисциплины.

Если в своем современном понимании учебная лексикография действительно представляет собой молодую отрасль языкознания, то с практической точки зрения она является чуть ли не самой древней его отраслью. При этом необходимо учитывать, что любой словарь представляет собой прежде всего дидактическое произведение. Лексикограф — автор словаря, так же как и автор учебника, выступает как посредник между обществом (коллективным носителем речи) и индивидуумом. Обращаясь за справкой к словарю, пользователь как бы обращается с вопросом к лексикографу и воспринимает его ответы как более или менее обязательные. Таким образом, назначение словаря любого типа — устранить расхождение между знанием индивидуальным и знанием всего коллектива. Анализ словарных дефиниций, характер отношения словарных высказываний к описываемому миру, свойственный словарю тип сообщения показывают, что любой словарь относится к дидактическому жанру.

Среди многочисленных функций словарей В. Г. Гак выделяет четыре основные группы.

1. Связанные с межъязыковым общением. Эти функции осуществляются преимущественно двуязычными словарями и разговорниками.

2. Связанные с обучением языку, разъяснением непонятных слов. Лексикографические материалы могут быть одноязычными или двуязычными, могут быть представлены глоссариями,

разговорниками, словарями. Обучение может касаться: а) языка культуры, отличающегося от разговорной речи (это может быть тот же язык прежней эпохи либо совсем другой язык, например, латинский, древнегреческий, старославянский в средневековой Европе, классический арабский в странах Востока, санскрит в Индии и т. п.); б) иностранных языков; в) родного языка.

3. Связанные с описанием и нормализацией родного языка. Эти функции осуществляются прежде всего толковыми словарями, а также специальными одноязычными словарями: орфографическими, трудностей и т. д.

4. Связанные с научным изучением языка. Выполнение этих функций вызвало к жизни различные типы одноязычных и двуязычных филологических словарей.

Если рассматривать проблему реализации всех перечисленных функций лексикографии в историческом аспекте, можно выделить следующие периоды ее развития.

1. Дословарный период (в Европе — до XV в.). Словарная продукция была представлена глоссариями, вокабулариями, ономастиконами и т. п., функциями которых было объяснение малопонятных слов языка культуры.

2. В ранний словарный период (в Европе — XV—XVII вв.) появились лексиконы и двуязычные словари, обслуживающие потребности межъязыкового общения, изучения языка культуры, обучения иностранным языкам.

3. Период развитой лексикографии (в Европе — с XVIII в.) характеризуется появлением словарей разных типов, в особенности толковых. Эти словари предназначены для выполнения всех четырех основных групп функций лексикографии.

Таким образом, учебная лексикография возникла на начальном этапе развития словарного дела. Даже названия многих словарей раннего периода содержат указания на их дидактическое назначение. Так, первый латинско-английский словарь назывался „Promptulorium parvulorum" — «Кладовая (слов) для детей». Ранние французские латинские словари назывались „Commentarius puerorum" — «Объяснение для юношества», „Seminarium et plantarium fructiferarum" — «Рассадник (знаний) для тех, кто может принести плоды» и т. д. Позднее эта учебная направленность отошла на второй план, уступив место составлению толковых, исторических, этимологических словарей.

В настоящее время, в результате постоянного укрепления международных связей, появления многих самостоятельных национальных государств, подъема культуры ранее колониальных народов, сопровождающего этот процесс, вновь происходит выдвижение на первый план тех функций лексикографии, которые связаны с межъязыковым общением и обучением языку. Все это привело к тому, что учебная лексикография осо-

бенно бурно развивается в два последних десятилетия (Гак, 1977).

Несмотря на это само понятие «учебная лексикография» по сей день, по существу, достаточно четко не определено. Существует целый ряд определений, из которых наиболее часто цитируется принадлежащее Л. А. Новикову: «Учебную лексикографию по сравнению с академической можно определить в целом как лексикографию меньших форм и большей обучающей направленности» (Бархударов, Новиков, 1971, с. 46). Это определение нуждается в существенных дополнениях и уточнениях с тем, чтобы оно могло охватить все виды описания семантических, синтаксических, стилистических, морфологических, дифференциальных, статистических и т. п. свойств слов, как представленных самостоятельными изданиями, так и являющихся составными частями других учебно-методических изданий. Анализируя литературу, специально посвященную лексике как объекту описания в учебных целях, В. В. Морковкин (Актуальные проблемы. . ., с. 31—35) выделяет пять главных аспектов такого описания.

1) Учебные словари. Проблематика анализа структуры, словника, типов учебных словарей разработана весьма подробно. Необходимо отметить, однако, что большинство исследователей концентрировало свои усилия, главным образом, на аналитических проблемах сходства и различия существующих словарей разных типов, способах семантизации лексики и т. д. В то же время проблема выработки критериев, создания процедур получения словарей того или иного типа является по сей день недостаточно разработанной.

2) Лексические минимумы. В настоящее время минимизация словарного состава языка обладает достаточно разработанной теоретической базой, существует и большое число лексических минимумов для разных типов и форм обучения. Однако для того чтобы полностью использовать все преимущества лексических минимумов в учебном процессе, их применение должно носить комплексный характер и охватывать все этапы обучения, что представляет собой не решенную в большинстве случаев проблему, так как разработка лексических минимумов в соответствии с научными принципами минимизации лексики является чрезвычайно трудоемким делом.

3) Лингвостатистические исследования, результаты которых часто представляются в виде частотных или распределительных словарей.

4) Учебные пособия по лексике лексикографического типа. Создано большое число таких пособий для разных языков, однако в теоретическом плане вопрос о таких пособиях практически не изучался.

5) Учебники. В учебниках лексика семантизируется в поурочных словарях, в сводном словаре, а также в специальных

лексических упражнениях. Лексический аспект языка чрезвычайно важен для любого учебника, однако теоретического обоснования сложившейся практики семантизации лексики в учебниках не имеется.

Как было показано выше, характер учебного словаря определяется прежде всего соответствующей методической направленностью, четкой ориентацией на тот или иной вид и этап обучения, целями и задачами обучения, родным языком обучаемых, наличием языка-посредника и т. п. Специфика учебного словаря заключается, прежде всего, в целенаправленном отборе языкового материала, проявляющемся уже на этапе формирования словника, т. е. составления исходного списка опорных единиц для последующего лексикографического описания. Словник словаря должен иметь минимально-достаточный характер, что может быть достигнуто путем строгого учета квантитативных параметров слов в их основных значениях, грамматических форм и синтаксических конструкций, путем отбора тематически наиболее существенных слов и словосочетаний. Необходимо также предусматривать проверку покрываемости словарем основных тем и речевых ситуаций, что, к сожалению, очень редко делается на практике.

Объем словника учебного словаря определяется, в первую очередь, его типом. Так, например, краткий толковый словарь должен отражать лексику такого носителя языка, который в данном языковом коллективе соответствует стандарту полноценного члена общества. Для решения этой задачи необходимо построить минимальную функционально-речевую модель определенного общества, включив в нее среднюю сумму культуры, научных и производственных знаний. Такой толковый словарь-справочник вполне может содержать 20 000—25 000 словарных статей. Однако такой словарь находится в пределах пассивного или даже активного словарного запаса отдельных индивидов и может быть отнесен к учебным словарям активного типа, но он не может служить надежным инструментом для перевода оригинальной литературы. Для этого необходим словарь пассивного типа существенно большего объема.

Тот факт, что, по существу, любой учебный словарь представляет собой результат строгого целенаправленного отбора, т. е. является лексическим минимумом в широком понимании, содержит в себе внутреннее диалектическое противоречие. С одной стороны, такой словарь должен в определенной степени покрывать все тексты заданного профиля и некоторой трудности, соотнесенные с данным этапом обучения, т. е. выполнять функции лексического минимума и быть не слишком конкретным, а с другой стороны, он должен содержать конкретный материал, необходимый для понимания любого текста данного рода. Эту задачу не может решить ни один словарь, созданный с использованием какого-либо одного принципа отбора (час-

тотного, тематического, ситуативного и т. п.). Слова образуют ряд классов, составляющих как бы разные «уровни» словаря.

Прежде всего это: а) структурные слова, образующие грамматический каркас языка и широко использующиеся в любом тексте; б) частотные знаменательные слова, использующиеся в большинстве текстов и поэтому слабо связанные с определенными темами, речевыми ситуациями; в) тематические, ситуативные слова, находящиеся в сильной связи с определенной темой или ситуацией. Эта категория слов не обладает высокой частотностью, употребляется в ограниченных речевых контекстах, но является необходимой для их реализации. Только введение в словарь единиц разных категорий (структурных, частотных, тематических, ситуативных) позволяет как обеспечить понимание широкого круга текстов заданного характера, так и соотнести учебный словарь с конкретным текстом, речевой ситуацией.

Специфика структурных компонентов учебного словаря. По сравнению с «большими» словарями структура учебных словарей характеризуется двумя основными особенностями: способами толкования (семантизации) лексических единиц, а также способами и приемами раскрытия парадигматических и синтагматических свойств и отношений этих единиц.

Семантизация лексических единиц является важнейшим структурным компонентом толкового (базисного) учебного словаря, в определенной степени она присутствует и в словарях других типов. В переводных учебных словарях семантизация, как правило, осуществляется при помощи переводных эквивалентов, в толковых словарях — определениями. Определение, являющееся важнейшим структурным элементом учебного словаря, *должно соответствовать требованиям краткости, наглядности*, сводимости к элементарным словам.

В толковых словарях, предназначенных для иностранцев, необходимо добиваться определения толкуемого понятия при *помощи строго ограниченного количества неопределяемых слов*, хорошо знакомых обучаемым. Эти слова образуют «минимальный словарь», при помощи которого определяются все прочие лексические единицы словаря. Так, например, в широко известном словаре М. Уэста (West, 1965) 24 000 слов объясняются при помощи около 1500 слов. В словаре основной лексики французского языка (Gougenheim, 1959) используются 1374 слова элементарного словаря и 55 слов-определителей. Известна попытка определения основной лексики русского языка при помощи ограниченного набора элементарных слов и понятий (Комараш, Феньвеш-Копяева, 1969).

Опыт составления учебных словарей показывает, что иногда определения, составленные при помощи «минимального словаря», страдают искусственностью и недостаточной определенностью. В подобных словарях встречается и «порочный круг» в

определениях, наличие которого считается недопустимым в словарях других типов. Однако такие недостатки объясняются не пороками принятого метода семантизации, а ошибками в работе составителей.

Особое значение для учебного словаря приобретает отражение синтагматических и парадигматических связей слов. Из практики известно, что даже знание лексики изучаемого языка не гарантирует правильного употребления слов в речи. Интерференция родного и иностранного языков ведет к образованию в речи обучаемых словосочетаний, не употребляемых в изучаемом языке. В этом плане большие филологические словари оказываются недостаточно эффективными для обучаемого-иностранца, так как они не содержат необходимых сведений об особенностях словоупотребления, в частности, о лексической сочетаемости слов. Поэтому на практике остро ощущается недостаток таких учебных словарей, в которых в качестве основного структурного компонента представлены наиболее устойчивые нормативные связи слов (см., напр.: Регинина и др., 1980; Дерибас, 1979).

Парадигматические отношения представлены в словарях парадигматического типа: тематических, синонимических, антонимических, в которых представлены парадигмы тематически связанных слов, а также в толковых словарях, где лексико-семантические варианты многозначного слова также представлены в определенных парадигмах.

Типология учебных словарей. Типологически учебные словари могут рассматриваться с нескольких точек зрения: 1) назначение словаря; 2) экстралингвистические параметры словаря; 3) внутрилингвистические параметры словаря.

По назначению учебные словари могут быть общими (предназначенными для всех категорий обучаемых) и специализированными (предназначенными для определенной категории обучаемых: школьников, студентов-филологов, студентов-нефилологов, преподавателей и т. д.).

Экстралингвистические параметры определяют специфику каждого словаря в той мере, в какой он близок к энциклопедическому словарю или содержит элементы энциклопедизма. В этом плане можно выделить такие единицы описания, как предметы материальной культуры, концепции духовной культуры, страноведческий компонент, а также такие аспекты описания, как характер отражения логических категорий в изучаемом языке, отражение в нем различных концептуальных «картин мира» на уровне индивидуального, группового и национального сознания. Экстралингвистические параметры являются определяющими для словарей имен собственных, поговорок, пословиц, крылатых выражений и т. п.

Типология словарей по внутрилингвистическим параметрам должна отражать типологию словарного состава языка. Сло-

варный состав любого языка насчитывает много сотен тысяч слов, поэтому для быстрого поиска нужного слова необходимо выработать стратегию, учитывающую системный характер лексики и позволяющую отыскивать необходимое слово не во всем словарном составе, а только в небольшой его части, например, в семантическом поле или лексико-семантической группе, на которую говорящего ориентирует ситуация, коммуникативное задание, логическая связность, ассоциативность мышления и т. п. В идеале теория учебной лексикографии должна разработать систему аспектных словарей (по аспектам лексической системы языка), которые, моделируя естественную стратегию поиска, сокращали бы время на поиск нужного слова.

Эта проблема затрагивает вопросы объема словаря, его организации, объема информации, содержащейся в словаре о каждой лексической единице, разработки системы всякого рода помет, отсылок и перекрестных ссылок, т. е. носит комплексный характер. Любой отдельно взятый учебный словарь способен решать лишь частную задачу и практически всегда используется как справочник. Задача лексикографа — сделать так, чтобы каждый словарь решал свою задачу оптимальным образом: при максимальном объеме лексического материала требовал бы минимального времени на поиск нужной информации.

Таким образом, по внутрилингвистическим параметрам учебные словари могут относиться к одной из следующих групп (отнюдь не взаимоисключающих).

1. Словари, организованные по уровням языка. К этой группе относятся словари морфем (аффиксов, корней, типов чередования — «внутренней флексии»), слов, словосочетаний (как в лексическом, так и в синтаксическом плане, например, словари глагольного управления) и словари предложений.

2. Словари, организованные по аспектам языка, а также по функционально-семантическим категориям. Сюда относятся лексические, словообразовательные, грамматические словари, а также комплексные или функционально-семантические словари.

3. Словари ономасиологические и семасиологические. Это деление вызвано двумя подходами к описанию языка — от формы к значению и от значения к форме, что, в свою очередь, связано с формированием двух основных речевых умений — рецептивного (т. е. умения понимать устную и письменную речь) и продуктивного (т. е. умения выражать свои мысли в устной и письменной форме). Оно выражает противопоставление словаря слушающего (читающего) словарю говорящего (пишущего). Соответственно можно говорить о словарях «пассивного» типа и о словарях «активного» типа. Основное же различие семасиологического и ономасиологического словарей заключается в том, что в первом словаре слова-значения распо-

лагаются по алфавиту, во втором — по классам, подклассам, группам и рядам, выражающим систему понятий. Примером ономасиологического словаря могут служить идеографический или тематический словари, словари аналогий, ситуаций и т. д.

4. Словари элементарной, основной, активной, пассивной, всей лексики языка; словари общеупотребительной и специальной лексики; словари нейтральной и стилистически маркированной лексики.

5. Словари одноязычные и двуязычные. Это противопоставление заключается в способе семантизации лексических единиц словаря: в первом случае она осуществляется с помощью единиц «минимального словаря», во втором — с помощью перевода на родной язык обучаемого или на язык-посредник. Это деление связано с различным решением проблемы выбора метаязыка описания лексических единиц словаря. И те и другие словари могут быть осложнены графическими иллюстрациями (рисунками, фотографиями, чертежами и схемами), что представляет собой включение в метаязык словаря коммуникативных средств иной семиотической природы.

6. Словари парадигматические и синтагматические. Словари первого типа представляют собой набор парадигм лексико-семантических вариантов слов (толковый словарь), синонимических рядов (словарь синонимов), антонимических противопоставлений (словарь антонимов), сходных по звуковому составу слов (словарь паронимов) и др., расположенных в алфавитном порядке. Особенностью словарей второго типа является раскрытие употребления слов в речи. Этот тип словарей представлен словарями сочетаемости, в которых слова также расположены в алфавитном порядке.

7. Словари комбинированные и некомбинированные. Одной из характерных особенностей учебных словарей является стремление отразить в одном словаре различные свойства и отношения слов в языке. На практике это приводит к тому, что чаще всего учебный словарь представляет собой комбинированный (комплексный) словарь. Большинство учебных словарей являются комбинациями нескольких параметров, представляющих собой типологические характеристики. Различные комбинации типологических характеристик создают «типологическую сетку» учебных словарей, многие ячейки которой еще не заполнены реальными словарями.

В ближайшее время предстоит создать систему учебных словарей, реализующих предсказанные типологией возможности, прежде всего тех, которые имеют наибольшее практическое значение.

Однако коллективный характер труда, трудоемкость, тесная связь научных аспектов с технологическими, организационными и финансовыми ставят ряд препятствий перед составителями словарей. Работа вручную, длительные сроки составления

картотеки и изготовления словаря обессмысливают учебную лексикографию, так как не позволяют избежать разрыва единого языкового сознания, морального старения словаря до его выхода в свет. Кроме того, необходимо предусматривать возможность использования одной картотеки для составления словарей для разных видов и форм обучения. Из этого положения есть только один выход — автоматизация лексикографических работ с использованием ЭВМ.

Литература

Актуальные проблемы учебной лексикографии. М., 1977.

Бархударов С. Г., Новиков Л. А. Каким должен быть учебный словарь? // Рус. язык за рубежом. 1971. № 3.

Вопросы учебной лексикографии. М., 1969

Гак В. Г. О некоторых закономерностях развития лексикографии (учебная и общая лексикография в историческом аспекте) // Актуальные проблемы учебной лексикографии. М., 1977.

Денисов П. Н. Очерки по русской лексикологии и учебной лексикографии. М., 1974.

Дерибас В. М. Устойчивые глагольно-именные словосочетания русского языка. М., 1979.

Комараш И., Феньвеш-Коняева Е. Лексический минимум русского языка. Будапешт, 1969.

Новиков Л. А. Учебная лексикография и ее задачи // Вопросы учебной лексикографии. М., 1969.

Проблемы учебной лексикографии. М., 1977.

Проблемы учебной лексикографии и обучения лексике. М., 1978.

Регинина К. Б., Тюрина Г. П., Широкова Л. И. Устойчивые словосочетания русского языка. М., 1980.

Теория и практика учебной лексикографии. М., 1988.

Учебные словари русского языка: Докл. сов. делегации. М., 1973.

Gougenheim G. Dictionnaire fondamental de la langue française. Paris, 1959.

West M. An International Reader's Dictionary. London, 1965.

АВТОМАТИЗАЦИЯ В ЛЕКСИКОГРАФИИ

Появление автоматических словарей (АС) и автоматических терминологических банков как лексикографических объектов особого класса было обусловлено несколькими причинами: с одной стороны, это рост возможностей ЭВМ, а с другой — осознание необходимости поиска путей повышения производительности труда лексикографа и переводчика.

Лексикография относится к одному из самых трудоемких видов человеческой деятельности. Создание того или иного словаря в завершенном виде требует порою десятилетий работы. Не случайно поэтому, что первые достижения по автоматизации лингвистических работ проникли прежде всего в лексикографию. Стало очевидно, что выдачу различных словников, указателей больших объемов целесообразнее, по возможности, передать машине; прежде всего, с помощью ЭВМ стали создаваться частотные и обратные словари.

Основное назначение АС — повышение производительности труда человека при работе с текстами, а также сбор, хранение и обработка различных единиц языка, подлежащих лексикографической интерпретации.

В целом АС является особым видом словаря прежде всего именно по функциональному критерию; он предназначен для использования в практических системах автоматической переработки текста (АПТ).

Обязательной частью любого АС является словник. Этот словник может быть входом в систему семантического описания (базу знаний), а может быть и единственной составляющей лингвистической базы.

Так, например, словари, создаваемые в системах информационного обслуживания, включают в себя дескрипторные словники, являющиеся входом в соответствующую базу знаний (тезаурус, систему фреймов). Словники систем машинного пере-

вода (МП) содержат лексические единицы (ЛЕ) с морфологическими, синтаксическими и базовыми семантическими характеристиками, необходимыми для осуществления лексического и синтаксического анализа текста. Кроме того, такие системы должны включать в себя АС описания значений, синтаксических и семантических правил. Эти словари составляют базы знаний систем МП.

Любая экспертная система также предполагает наличие словника, обеспечивающего доступ к базе знаний.

В системах устного ввода словник включает в себя написание словоформы, транскрипцию.

Словник одноязычного АС толкового типа полностью определяется теми принципами, которые заложены в проект такого словаря.

Автоматические переводные словари (АПС) предназначены в основном для обработки научно-технических текстов, поэтому бо́льшая часть состава АПС представлена терминологической лексикой. Для повышения эффективности работы крупного АПС в него необходимо включать всю терминологическую лексику по тематике словаря, наиболее употребительную лексику из сопредельных тематических областей, общенаучную, общетехническую, а также наиболее употребительную общую лексику. Общая лексика нужна, в первую очередь, техническим специалистам, которые обычно хорошо знают лексику собственной специальности, но недостаточно хорошо знают иностранный язык в целом.

АС предназначены для обработки всех видов научно-технических текстов, поэтому они должны включать в себя также термины как стандартизованные и рекомендуемые, так и не рекомендуемые, как устоявшиеся, так и новые.

При определении структуры АС должны быть решены три основные задачи: 1) определение формы заглавной единицы в словарной статье; 2) определение комплекса информации, составляющей описание ЛЕ; 3) выбор оптимального способа организации словарной статьи в базе.

Существенное значение для АС имеет выбор основной единицы его словника. Для крупных словарей общей лексики основной единицей словника является слово. Словосочетания даются обычно в составе статей на отдельные слова — либо в качестве иллюстрации значений, либо как пример сочетаемости заглавных ЛЕ.

С точки зрения основной единицы своего словника АПС примыкают к традиционным терминологическим словарям. До 80% их лексических массивов составляют словосочетания. Крупные АПС в своем большинстве фактически являются терминологическими банками и направлены на решение тех же задач, что и традиционные терминологические словари. Верхняя граница словосочетаний в этих АПС еще больше размыта даже

по сравнению с традиционными терминологическими словарями. В крупные АПС нередко включаются в качестве самостоятельных статей отдельные предложения и даже фрагменты.

Подобный подход к комплектованию словника переводного словаря в традиционной лексикографии не используется. Однако в АС такое решение в определенных условиях оказывается весьма удобным. Дело в том, что АПС показали себя эффективным инструментом полуавтоматического перевода стандартизованных текстов, например, товаросопроводительной документации или юридических документов, в которых определенные части регулярно повторяются без существенных изменений.

Современные промышленные АПС создаются на базе мощных ЭВМ, имеющих огромные ресурсы памяти, которые позволяют успешно работать с массивами в сотни тысяч словосочетаний. Вместе с тем нестрогий подход к верхней границе словосочетаний приводит к тому, что память ЭВМ используется недостаточно эффективно, поэтому, несмотря на большие и постоянно растущие возможности современных ЭВМ, следует избегать перенасыщения АПС свободными словосочетаниями. Такие единицы текста целесообразно использовать в качестве входных единиц словаря только в тех специфических случаях, когда это дает безусловный выигрыш в качестве и скорости перевода.

Как в общей лексикографии, так и в двуязычной (или многоязычной) выбор словарной единицы определяется и спецификой слово- и формообразования конкретных языков.

Схема взаимодействия «текст — АС» вполне оправдывает себя с точки зрения построения быстродействующих систем переработки текстов на языках изолирующего, аналитического и иногда флективно-аналитического типа. В этом случае на уровне досинтаксического анализа мы можем получить всю информацию, приписанную данному слову в словаре, без привлечения алгоритмов морфологического анализа. Так, например, английские словоформы was, were, said, received независимо от регулярного или нерегулярного способа образования могут непосредственно храниться в АС, и отождествление их текстовых «близнецов» происходит с максимальной скоростью.

Иная ситуация возникает при обращении к синтетическим флективным и агглютинирующим языкам, для которых использование словоформ в качестве базовых единиц АС связано с резким увеличением объема словаря, что обесценивает быстроту отождествления словоупотребления текста со словоформой из АС. Так, русское существительное, например, имеет до 12 форм, глагол при учете видовой корреляции образует до 188 форм. В тюркских языках именная ЛЕ теоретически может иметь бесконечное число форм.

В связи с этим возникает необходимость компрессии таких

парадигм. Эта задача может быть решена двумя способами: 1) лингвистическим (алгоритмическим); 2) техническим.

При лингвистическом подходе основное внимание уделяется формированию единицы словаря — установлению эталона, образца, наиболее подходящего инварианта. Для синтетических, флективных и агглютинативных языков используются различные приемы компрессии, среди которых наиболее эффективным методом является организация АС в виде словаря основ. В этих АС обычно выделяют два вида единиц: собственно основы и так называемые особые формы, т. е. такие словоформы, которые образованы от основ чередованием или супплетивно; они включаются в АС наравне с основами.

В то же время для ряда языков оказывается недостаточно введения одной основы для парадигмы даже при использовании особых форм. Так, во французском языке при таком подходе у многих глаголов выделяется от двух до пяти основ. Например, для глагола vouloir выбраны три основы — vouill, voud, voul и три особые формы — veut, veux, veulent.

Для моделирования лексической системы русского языка, имеющего развитую флексию, используются словари основ. Разумеется, компоновка этих словарей может быть различна. Так, например, если русский АС «МАРС», разработанный в Лаборатории инженерной лингвистики РГПУ им. А. И. Герцена и включенный в качестве ядра в систему СИЛОД, организован как словарь основ и типовых парадигм, то словарное обеспечение системы GETA предусматривает использование семи русских словарей: два словаря в этой системе содержат основы, два — неизменяемые слова и особые формы, два — суффиксы и окончания, и наконец, один словарь включает в себя машинные обороты.

Структура словарной статьи в разных АПС весьма различна и колеблется от 10 до 99 зон, в которых фиксируется разнотипная лингвистическая и экстралингвистическая информация. Разные АПС включают в свои словарные статьи в различных сочетаниях следующие сведения: заглавное слово, один или несколько его иноязычных эквивалентов, тематическую принадлежность заглавного слова и его иноязычных эквивалентов, грамматическую информацию, неформализованные толкования или стандартизованные дефиниции, контексты, лексическую сочетаемость, семантические иерархические связи заглавного слова, код качества (так называемый индекс надежности), сокращения, стилистическую окрашенность, лингвогеографические ограничения, ключевые слова (если заглавная лексическая единица является словосочетанием), источники, фамилию автора словарной статьи, дату поступления, различную служебную информацию.

Соотношение семантики, синтаксиса и морфологии в струк-

туре словарной статьи (СлСт) определяется особенностями ее организации.

Каждая лексическая единица, входящая в базу данных, получает описание на морфологическом, синтаксическом, семантическом и функциональном уровнях в виде соответствующего набора характеристик. Синтаксические и морфологические характеристики необходимы для определения значения текстовой словоформы и установления ее синтаксической функции в предложении. Семантические характеристики описывают значение ЛЕ, а функциональные — потенциальную роль единицы в общей структуре документа.

Для размещения только что рассмотренных характеристик ЛЕ в современных системах АПТ используются словарные статьи трех видов.

1. Словарные статьи единой жесткой структуры, имеющие фиксированную длину. Эта организация предусматривает включение в статью всех параметров описания, характеризующих данный класс ЛЕ. Такой подход может распространяться на все описание ЛЕ, включая ее семантические и функциональные параметры, а может ограничиваться синтаксической информацией. Во втором случае аналогично организованная семантическая информация вводится в лингвистическую информационную базу (ЛИБ) отдельным словарем.

При такой структуре заранее определены все типы информации, включаемые в статью, и фиксированы их границы в байтах. При отсутствии информации конкретного типа соответствующие поля остаются свободными. Примером такой жестко организованной СлСт является статья системы SYSTRAN общим объемом в 255 байтов. АС этой системы организован как словарь основ, каждая СлСт должна содержать: Y — описание ЛЕ с указанием способа ее обработки при вводе в ЛИБ (добавление, коррекция, уничтожение); A — описание базовой машинной основы; B—морфологическую информацию о падежных окончаниях, формах множественного числа и т. п.; C — синтаксические коды, включающие в себя описание управления; D — семантические коды; E — коды переводов предлогов, эта позиция заполняется только для предлогов и пустует для всех остальных частей речи; F — основные переводы, включающие в себя описание особенностей эквивалентов.

Такой метод организации позволяет достаточно просто пополнять и корректировать информацию в СлСт в том случае, когда для нее заранее определена позиция. Однако введение любой новой, ранее не предусмотренной информации влечет за собой коренную перестройку СлСт, самого словаря и базы. С этим связана громоздкость такой структуры — в статье должны фиксироваться все потенциально возможные характеристики ЛЕ независимо от того, разработаны ли модули системы АПТ, которые их используют. Большинство позиций в такой

статье, особенно в промышленных системах, пустует. В то же время каждая СлСт описываемого типа занимает большой объем памяти ЭВМ. Это, с одной стороны, вызывает сложности размещения АС, с другой — нарушает запрет на дублирование информации.

2. Словарная статья с плавающей правой границей, в которой описание задается по принципу признак — значение. Формирование описания в такой статье также происходит по единой схеме, которая предполагает разделение всей информации о ЛЕ на зоны, число которых определяется в зависимости от конкретного типа сведений. Эти зоны описывают ЛЕ с позиций класса слов, лексико-грамматических, морфологических, семантико-синтаксических, формальных, функциональных и т. д. характеристик. В каждой зоне выделяются признаки, количество которых зависит от класса, к которому принадлежит ЛЕ. Так, например, класс детерминативов, выполняющих функцию актуализации существительного, характеризуется следующим набором признаков, описывающих конкретные единицы этого класса: подкласс слов, число, род, позиция и т. д.

«Плавающая правая граница» существует только на этапе предмашинного заполнения словарной статьи, так как при вводе в память ЭВМ это описание должно преобразовываться в такую же жесткую структуру, как и в первом случае.

3. Словарная статья с иерархической структурой, предусматривающая вынесение части информации в отдельные файлы (массивы). В этом случае АС как бы делится минимум на две части: список словарных входов с приписанной обязательной информацией и словарь значений.

При этом в соответствии со стратификационным подходом предварительно проводится иерархизация всех характеристик ЛЕ с учетом их использования при АПТ: выделяются верхние и нижние уровни иерархии. Характеристики верхних уровней, непосредственно используемые при любых видах АПТ, вводятся непосредственно в структуру статьи, все остальные характеристики, как правило — семантические, используемые особыми процедурами АПТ, выносятся в отдельный файл или файлы. В самой словарной статье в этом случае указываются только отсылки к соответствующим файлам.

В базовом варианте в структуру СлСт входного словаря включается: сама заглавная ЛЕ (основа, словоформа или словосочетание); лексико-синтаксический код, зависящий от типологических особенностей входного языка, грамматики и реализованных в системе алгоритмов синтаксического анализа; код подъязыка, определяющий особенности функционирования слова в разных предметных областях; информация о переводе.

Верхний уровень описания (словник) включает в себя саму ЛЕ и такие ее морфологические характеристики, которые необходимы для проведения первичного синтаксического (синтагма-

тического) анализа предложения. Так, например, в вопросно-ответной системе верхний уровень описания может состоять из самих ЛЕ, каждой из которых приписана синтаксическая и морфологическая информация. Этот уровень называется определением и состоит из двух частей: собственно определяемого и пары. Определяемое задается основой слова в случае регулярного словообразования или словоформой при нерегулярном формообразовании, например:

Определяемое	Пара
BOOK	(N — S)
OX	(N — IRR)
OXEN	(N — (OX(N — PL)))
INCLUDE	(V — S — D)
WRITE	(V — IRR)

Пара состоит из описания лексической категории — класса слов (всего выделено 17 классов) и ее значения, описываемого морфологическим кодом (ср. BOOK) и списком характеристик корня (ср. OXEN, WRITE), включающим в себя описание нерегулярного формообразования. Описание нижнего уровня приписывается в этой системе глаголам. В качестве отсылки к файлу семантических описаний используется признак глагола (V) и сама ЛЕ.

Более детальное синтаксическое описание при использовании иерархической структуры создается в отдельном АС так, что его статьи связаны с СлСт лексического уровня с минимальным дублированием информации. Например, СлСт английских лексических единиц president, minister, envoy и т. п. имеют одинаковое синтаксическое и семантическое описание, и, следовательно, в их статьях используется отсылка на одно и то же описание. Аналогичным образом формируются отсылки к АС семантических описаний.

При такой структуре модульный и стратификационный подходы согласуются с принципом отсутствия дублирования информации.

Рассмотрим некоторые особенности лингвистического описания, которое может быть введено в структуру словарной статьи также на примере АПС, разработанного во Всесоюзном центре переводов (ВЦП).

Словарная статья разделена на 13 зон. Каждая зона содержит строго определенную информацию и имеет свои правила заполнения.

Зона 1. Зона заголовка является ключевой зоной словарной статьи, по которой осуществляется вход в словарную статью, производится ее формирование и поиск. В качестве заголовка могут выступать морфема, отдельное слово, словосочетание в своей канонической форме или фрагмент текста. Для полных омонимов создаются разные словарные статьи, которые сопровождаются порядковыми номерами.

З о н а 2. Зона лексико-грамматического класса (ЛГК) заголовка следует за заголовком. Одноэлементные лексические единицы подразделяются на группы в зависимости от своей частеречной принадлежности, а внутри этих групп классифицируются *по ряду категориальных признаков. Словосочетания* делятся на типы в зависимости от функции, выполняемой ими в предложении.

З о н а 2а. Зона ЛГК русского переводного эквивалента (РПЭ). Непосредственно за РПЭ записывается индекс его ЛГК. Его включение в исходную словарную статью необходимо для перевертывания словарной статьи, т. е. автоматического создания русско-иностранной словарной статьи.

З о н а 3. Зона морфо-грамматической информации (МГИ) служит для построения всей словоизменительной парадигмы заглавной единицы и, таким образом, позволяет осуществлять по*иск любой текстовой формы.*

Внутри зоны МГИ предусмотрена подзона сокращений, куда записываются все варианты сокращений заглавной лексической единицы (как для отдельного слова, так и для словосочетаний). Например: АНГЛ. ЯЗ. FUNCTION(N)F, FN, FUN, FUNC & FUNCTIONS... ФР. ЯЗ. ARRE<T SUR INSTRUCTION (NP-M) AI&...

З о н а 3а. Зона сокращений в РПЭ. В словарных статьях, где РПЭ имеют сокращения, введена зона 3а, которая следует сразу за зоной 2а (зоной ЛГК РПЭ).

З о н а 4. Зона тематической рубрики и стиля. В этой зоне приводятся тематические и стилистические пометы, уточняющие область и условия употребления заглавной единицы в данном значении. В этой же зоне может быть дано указание на язык-источник для заимствованной лексики. На основе этих помет автоматически по типам помет могут быть сформированы монотематические словари или списки слов, имеющих определенные стилистические особенности.

З о н а 5. Зона индекса надежности. Каждому РПЭ предшествует индекс надежности, отражающий степень общепринятости данного переводного эквивалента. Этот индекс определяется надежностью лексикографического источника, из которого взята лексическая единица. Для обозначения надежности РПЭ введены следующие индексы:

А — переводной эквивалент зафиксирован в каком-либо стандарте (ГОСТ, ОСТ, международный стандарт) либо рекомендован к употреблению авторитетными международными орг*анизациями, например:* АНГЛ. ЯЗ. COMPUTER(N) COMPUTERS ВЧТ*А 1. ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНАЯ МАШИНА (ИМ. СОЧ.-Ж) ЭВМ S ...

Б — переводной эквивалент зафиксирован в надежных апробированных словарях, например: НЕМ. ЯЗ. FREIPROGRAM-

MIERTE RECHENANLAGE (NP-F) ВЧТ Б 1. ВМ С ГИБКОЙ ПРОГРАММОЙ (ИМ. СОЧ.-АББ) 0007...

В — переводной эквивалент зафиксирован в тетрадях новых терминов, экспресс-выпусках терминологии и т. п., например, АНГЛ. ЯЗ. WHISKER(N) WHISKERS ЭЛЕКТРОНИКА В 1. ЭЛЕКТРОННЫЙ ВЫВОД (ИМ. СОЧ.-М)... 0213 ЭЛЕКТРОНИКА В 1. «УСИК» (М) ... 0213.

Г — переводной эквивалент обнаружен в переводных журналах, неопубликованных переводах или создан составителем словарной статьи.

Зона 6. Зона РПЭ. РПЭ представляют собой отдельные слова или словосочетания, записанные в канонической форме. Все РПЭ снабжаются порядковыми номерами. Синонимичные РПЭ имеют одинаковые порядковые номера, а переводные эквиваленты, выражающие различные значения заглавной лексической единицы, — разные номера.

В традиционной лексикографии существует понятие безэквивалентной лексики. Сюда относятся лексические единицы (ЛЕ), не имеющие перевода на другой язык. Это или общеупотребительная лексика, обозначающая понятия, отсутствующие в других странах (например, *рушник, варенец* и т. п.) или новые термины, для которых еще нет соответствующего термина в другом языке. В АПС ВЦП входит безэквивалентная лексика второго типа. В словарных статьях таких ЛЕ в зоне РПЭ ставится знак (@———), указывающий на отсутствие переводного эквивалента, но обязательно заполняется зона толкования. При выдаче на печать словарная статья с безэквивалентной лексикой имеет такой вид, например: АНГЛ. ЯЗ. FRAME(N) ВЧТ* Б 5. (@——) ТО: ПРЕДСТАВЛЕНИЕ В НАГЛЯДНОЙ ФОРМЕ ТЕКУЩИХ ВЫБОРОЧНЫХ ДАННЫХ ... 0037 ВЧТ* Б 6. (@——) ТО: ОДНА ИЛИ НЕСКОЛЬКО СТРОК БИТ, РАСПОЛОЖЕННЫЕ ПОПЕРЕК ЛЕНТЫ ... 0042 ...

Если со временем новому термину будет найден соответствующий этому новому значению переводной эквивалент, его можно дописать в зону РПЭ, и тогда словарная статья будет иметь стандартный вид.

Зона 7. Зона пояснительных помет. Пояснительные пометы служат для уточнения значений ЛЕ и описания правил их употребления в тексте. Эта информация необходима для правильного выбора переводного эквивалента. При «перевертывании» словарной статьи, т. е. при автоматическом формировании русско-иностранного варианта, пояснительные пометы сохраняют свое место при переводном эквиваленте, в качестве которого уже будет выступать иноязычная ЛЕ. Пояснительная помета может быть выражена в терминах семантических требований, предъявляемых данной ЛЕ к своему потенциальному главному актанту (т. е. указывается, представители каких се-

мантических классов могут заполнять позицию главного актанта при соответствующей русской лексеме), например: АНГЛ. ЯЗ. ADJUST(VT) ВЧТ*Б 1. ПОДСТРАИВАТЬ (ПЕР. НОВ) (УСТРОЙСТВО) ——— 0011.

Семантические пометы могут также задавать семантический признак самой заглавной ЛЕ и ее переводного эквивалента, например, АНГЛ. ЯЗ. LISTING(N) ВЧТ*Б 1. РАСПЕЧАТКА (Ж, ЕД) (ДЕЙСТВИЕ) ——— 0012 ВЧТ*Б 2. РАСПЕЧАТКА (Ж) (РЕЗУЛЬТАТ ДЕЙСТВИЯ) ——— 0012 ...

Лексические пометы описывают лексическую сочетаемость РПЭ и заглавной ЛЕ в терминах конкретных лексем, например: АНГЛ. ЯЗ. RUN(N) ВЧТ*Б 2. ПРОГОН (М) (ПРОГРАММЫ) ——— 0012 ВЧТ*Б 3. ЭКСПЛУАТАЦИЯ (Ж, ЕД) (МАШИНЫ) ——— 0012 ВЧТ*Б 5. ОТРЕЗОК (М) (ВРЕМЕНИ) ——— 0011.

Синтаксические пометы описывают синтаксическую модель управления РПЭ в терминах синтаксических позиций и особенности грамматического поведения заглавной ЛЕ в данном значении, например: АНГЛ. ЯЗ. ACCESS(N) ВЧТ*Б 1. ДОСТУП (М) (К ЧЕМУ-ЛИБО) ——— 0011 ОБР. ДАННЫХ Б 2. ВЫБОРКА (Ж) (ИЗ ЧЕГО-ЛИБО) ——— 0099 ...

Зона 8. Зона толкований. Для описания значений терминологических единиц могут быть использованы толкования. Толкование является надежным средством разграничения омонимии и многозначности в АПС.

Зона 9. Зона примеров употребления (контекстов). РПЭ могут сопровождаться примерами употребления заглавной ЛЕ в соответствующих значениях.

Зона 10. Зона фразеологии. Выделяется подзона, в которой записывается информация, характеризующая заглавное слово или словосочетание с точки зрения его положения в лексической системе языка, например: АНГЛ. ЯЗ. (А) ФЗ, SYN: LARGE, GREAT ANT: LITTLE, SMALL.

Довольно часто встречаются ЛЕ, имеющие несколько вариантов написания. В этом случае составляется столько СлСт, сколько существует вариантов написания, при этом в каждой СлСт в зоне 10а в качестве синонимов перечисляются все остальные варианты, например: АНГЛ. ЯЗ. ANKER (N) ФЗ: SYN ANKOR ANKOR(N) ФЗ: SYN ANKER.

Зона 11. Фамилия, имя и отчество составителя или редактора СлСт.

Зона 12. Название источника. Для каждого переводного эквивалента в СлСт указывается его источник.

Зона 13. Дата составления или редактирования СлСт.

Для снятия многозначности синтаксической и семантической интерпретации предложения и задания семантики нулевого уровня нередко используются автоматические словари оборотов (АСО), позволяющие выделять в тексте и переводить последо-

вательности, значение которых не равно сумме значений составляющих их элементов. Словари оборотов вводятся в структуру ЛИБ как отдельные словари или в составе единого АС. Функция этих словарей определяется тем, что большинство современных технических реалий передается не отдельными словами, а словосочетаниями.

Межъязыковая идиоматичность, затрудняющая получение корректного МП и распознавание смысла текста даже с использованием алгоритмов семантического анализа, также определяет необходимость введения словаря оборотов.

При построении словаря оборотов отбор сегментов производится с учетом статистического критерия: в АСО включаются словосочетания, встретившиеся с достаточно высокой частотой в различных частотных словарях словосочетаний и манифестирующие научно-технические реалии,— а также критерия переводимости: в АСО включаются сегменты входного текста, структура перевода которых не адекватна входной структуре.

АС семантических описаний как самостоятельный модуль словаря может быть организован несколькими способами: а) в виде набора семантических признаков или множителей, скомпонованного по определенным правилам; б) в виде расширенной или ассоциативной семантической сети.

Как правило, крупные АПС включают в себя лексику нескольких тематических областей, поэтому проблема пересечения тематических областей для АПС стоит весьма остро.

Создатели промышленных АПС уделяют решению этой задачи серьезное внимание и разрабатывают для этой цели специальные системы тематической классификации лексики. В основу тематической классификации лексики в АПС ВЦП положен рубрикатор Государственной автоматизированной системы научно-технической информации.

Поскольку практические системы МП являются прагматически ориентированными, то АС МП также жестко ориентированы на конкретные предметные области (ПО), что позволяет снимать на уровне лексического анализа многозначность ЛЕ и стандартизировать перевод терминологии.

Ориентация на конкретную ПО обязательна для АС любых систем, в случае создания многоаспектных систем код принадлежности к конкретной ПО вводится в структуру СлСт и является основанием для выбора перевода при лексическом анализе. Код ПО является единственной квазистилистической пометой, используемой в АС систем МП.

Основополагающим принципом при построении многоязычного АПС является способ организации связей между разноязычными ЛЕ в данном словаре. В промышленных АПС вне зависимости от количества языков используются три основных способа организации межлексических связей.

1. Каждая ЛЕ описывается независимо для каждого из

языков и только средствами данного языка (сведения о заголовке, грамматическая информация, толкование, все виды лингвистической и другой информации).

2. Создается общая СлСт для двух разноязычных ЛЕ, являющихся переводными эквивалентами друг друга.

3. Создается единая СлСт, объединяющая переводные эквиваленты на всех языках, включенных в словарь.

Под обратимостью словаря понимается возможность произвольно, по желанию пользователя, менять местами входной и выходной языки.

В многоязычном АПС проблему обратимости необходимо рассматривать на двух уровнях — на уровне языков и на уровне отдельных ЛЕ. Обратимость многоязычного АПС на уровне языков (языковая обратимость многоязычного словаря) означает, что языки, входящие в состав этого словаря, могут выступать как в качестве языков запроса, так и в качестве языков ответа. Под обратимостью на уровне ЛЕ (лексической обратимостью) понимается возможность использовать одну и ту же ЛЕ в качестве запроса и в качестве ответа.

Все крупные АПС являются частично или полностью обратимыми. Необратимость АПС является его крупным недостатком. По этой характеристике АПС четко противостоят крупным переводным словарям в традиционной книжной форме и системам МП, но имеют много общего с традиционными переводными терминологическими словарями.

Свойство обратимости АПС позволяет использовать его для автоматизации некоторых видов работ в традиционной переводной лексикографии. Основной резерв экономии времени при составлении переводного словаря лежит в возможности автоматически менять местами входной и выходной языки словарных статей.

Многоязычие является, пожалуй, самой броской отличительной чертой всех крупных АПС. Так, автоматический словарь LEXIS, который эксплуатируется уже более 20 лет, в настоящее время включает в себя лексические массивы немецкого, английского, французского, русского, итальянского, датского и португальского языков. АПС EURODICAUTOM охватывает шесть языков: английский, датский, итальянский, немецкий, нидерландский и французский. Словарь TFAM в настоящее время содержит около 2 млн. входных единиц на девяти языках. Автоматический словарь ВЦП и терминологический банк Дрезденского технического университета включают в себя четыре языка: английский, немецкий, французский и русский.

Под динамичностью АС понимается возможность оперативно и без существенных трудностей корректировать словник словаря, а также состав и объем информации в СлСт. По этой характеристике словари в традиционной книжной форме являются статичными. В отличие от традиционных словарей АС

обладают свойством динамичности. Отбор лексики не прекращается перед составлением словаря, а продолжается в течение всего времени его функционирования. Постоянно происходит исключение оказавшихся ненужными слов и включение новой лексики, коррекция и пополнение информации в любой зоне СлСт. Переменность состава и способность воспринимать изменения в ходе фунционирования являются принципиальными характеристиками, отличающими машинные словари от обычных.

В последние годы особенно широкое распространение приобрело построение с помощью ЭВМ словарей-конкордансов. В настоящее время мы располагаем уже фундаментальными конкордансами к произведениям Шекспира, Сенеки, Ронсара, Пушкина и др.

Обычно словоуказатель представляет собой алфавитно упорядоченный список словоформ (или слов) с указанием всех вхождений того или иного слова в заданный массив текстов. Конкорданс включает в себя те же данные, что и словоуказатель, но, кроме того, в нем для каждого слова приведен контекст, иллюстрирующий данное словоупотребление. В отличие от толковых словарей в конкордансах отсутствуют определения значений и соответственно распределение иллюстративных примеров по значениям.

Конкордансы, построенные при помощи ЭВМ, освобождают лексикографа от множества трудоемких операций по выборке, созданию словников, формированию массива картотеки цитат. При соответствующей ручной доработке любой конкорданс может стать основой, фундаментом толковых словарей разных типов.

Создание словарей описываемого типа может осуществляться как вручную, так и при помощи ЭВМ. В последнее время появился ряд статей, убедительно демонстрирующих преимущества автоматического составления словоуказателей и конкордансов. Анализ работ по составлению конкордансов свидетельствует о том, что наиболее актуальным в настоящее время является проектирование и создание специальных целостных автоматизированных словарно-справочных служб конкордансного типа, построенных по принципу «запрос — ответ». В результате одноразового ввода в ЭВМ текста тех или иных источников потребитель впоследствии получает широкие возможности неоднократного обращения к данной автоматизированной системе с самыми разными запросами. Подобные системы в виде так называемых словарных банков данных не только обеспечивают предельную точность, надежность и проверяемость хранения всей информации, содержащейся в тексте, но и, что, вероятно, самое главное, дают возможность получения ответов обо всех употреблениях этого слова, обо всех контекстах на каждое слово, на каждую словоформу, обо всех

типах вариантов как по отдельному источнику, так и по той или иной совокупности источников. О том, сколь необходимо в ряде случаев приведение в словаре полного или хотя бы максимально возможного перечня иллюстративных примеров, убедительно писал еще Л. В. Щерба.

Такие автоматизированные словарно-справочные службы могут быть ориентированы на тексты памятников древней письменности, на произведения классической литературы отдельных авторов, на отражение специальной терминологии, на тексты на разных языках.

Задача построения таких автоматизированных лингвистических баз данных, выполняющих словарно-справочные функции, чрезвычайно актуальна и, очевидно, будет решена в течение ближайшего времени.

Таким образом, создание как автоматизированных словарно-справочных систем, так и собственно конкордансов представляет собой актуальную комплексную общефилологическую проблему, которая может быть решена только в тесном содружестве лингвистов, литературоведов, математиков и специалистов по прикладной лингвистике.

Механизация и автоматизация терминологической работы в сочетании с традиционными методами обработки текстов научной и технической литературы и документации обеспечили создание терминологических банков — систем, состоящих из ряда функциональных элементов. Задачей этих систем является сбор, хранение и распространение терминологических данных в наиболее короткие сроки с наименьшим поисковым шумом и при наиболее низкой стоимости.

В последнее время как у нас в стране, так и в ряде других стран (Франция, Швеция, Бельгия, ФРГ, Канада) развернулась широкая работа по созданию сводных терминологических банков данных (ТБД) по различным отраслям знания, на очереди проблема обмена данными и телесвязи между банками.

Так, например, основой автоматизированной системы информационно-терминологического обслуживания в России является Справочный банк терминов, который содержит терминологию различных областей науки и техники. Источниками для данного банка являются государственные и отраслевые стандарты на научно-технические термины и определения, международные стандарты на терминологию, сборники рекомендуемых терминов.

В принципе ТБД может быть организован, с одной стороны, как документально-фактографическая информационная система (АДФИПС), выдающая сведения о понятиях и терминах, их обозначающих, а с другой стороны, как автоматизированная словарно-справочная информационная система, ориентированная на специальный текст.

ТБД следует рассматривать в двух аспектах. Прежде всего, это определенный тип автоматизированной информационной системы. Первый аспект связан с рассмотрением ТБД с позиций теории научно-технической лексикографии, и здесь при его проектировании встают те же вопросы, что и при разработке проекта обычного терминологического словаря (выбор и ограничение предметной области знания, отношение к истории науки, отбор языков, источников, методов формирования словника и т. д.).

Именно в научно-технической лексикографии, и прежде всего при создании терминологических словарей, автоматизация может быть использована не только для обработки источников сбора материалов и его сортировки, но и при написании СлСг. Термины образуют довольно четкие по семантике группы типа названий приборов, станков, приспособлений, процессов, которые делятся на большое количество конкретных видов. Все это позволяет заранее задать определенную модель параметров описания семантики терминов данного типа. Так, например, при описании многих приборов должны учитываться такие параметры, как назначение, место установки, состав деталей, принцип и зона действия, при описании биологического вида — указание на семейство, род, морфологические и биологические особенности, ареал, хозяйственное назначение. Параметры делятся на общие и частные. В каждой области знания свои типы свойств, процессов, локализаций, сфер действия и т. д. Параметры описания можно группировать по их позиции в определении термина, в зависимости от характера науки, ориентации словаря.

Но ТБД — это одновременно и автоматизированная система документально-фактографического типа. И здесь встают традиционные вопросы лингвистического и математического обеспечения автоматизированных систем (разработка словарей, методик индексирования документов, программ, диалога пользователя с ЭВМ). При успешном решении всех вопросов в идеале ТБД — отраслевого и межотраслевого типа — это достаточно глубокая, семантически сильная документально-фактографическая словарно-справочная диалоговая система.

Организация сводных документально-фактографических ТБД окажется в перспективе едва ли не основным звеном в работах по созданию толковых и переводных словарей, справочников, рубрикаторов, классификаторов, по стандартизации терминологии, обслуживанию терминологических комиссий, практики научно-технического перевода. Успехи и достижения современной вычислительной техники, прикладной лингвистики и информатики указывают на возможные экономные и эффективные пути создания таких автоматизированных систем.

В ТБД, в принципе, могут быть введены самые разные сведения о словах-терминах, в том числе и такие, как указания на

часть речи, происхождение слова, употребительность в текстах. ТБД, выдавая сведения о термине как о слове, обо всех его текстовых употреблениях, в силу самой природы термина представляет сведения о понятии.

Литература

Лингвистическая концепция терминологического банка данных машинного фонда русского языка: (Проект) / Под ред. А. С. Герда. М., 1989.

Марчук Ю. Н. Проблемы машинного перевода. М., 1983. 232 с.

Марчук Ю. Н. Машинный перевод как информационная и технологическая реальность. М., 1988.

Перстнев И. П., Волкова И. Н. Банки терминологических данных. М., 1987.

Попов Э. В. Общение с ЭВМ на естественном языке. М., 1982. 360 с.

Ritchie G. D., Hanna F. K. Semantic Networks — a General Definition and a Survey // Information Technology. 1983. Vol. 2, N 4. P. 187—231.

Vauquios B. Aspects of Mechanical Translation in 1979: Conference for Japan IBM Scientific Program. Grenoble, 1979. 50 p.

Vauquios B. La traduction automatique a Grenoble // Documents de linguistique quantitative. N 24. Paris, 1975. 184 p.

Wang W. S.-L., Chan S. W., Ts'ou B. K. Chinese Linguistics and Computer // Linguistics. 1973. 118. P. 89—117.

ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ

Процесс развития познания и производительных сил в последние десятилетия привел к возникновению новой науки — информатики.

Информатика, с одной стороны, рассматривается как наука о наиболее общих закономерностях организации и преобразования информации. С другой стороны, сегодня это новая отрасль общественного производства. Зарождение и развитие информатики прежде всего связано с развитием новых средств обработки информации, в первую очередь, вычислительной техники и средств связи. Новые методы информационного обеспечения научно-технического прогресса характеризуются широким применением автоматизированных систем. Системы, связанные с электронной обработкой данных, решают сегодня самые разные задачи: расчетные, технологические, учетно-статистические, издательские, а также чисто информационные, т. е. задачи хранения и поиска информации.

Одно из важных следствий научно-технического прогресса — небывалый рост объемов информации. Рост числа научно-технических изданий и публикаций резко обострил проблему отыскания нужных сведений. По образному выражению акад. С. И. Вавилова, современный читатель находится перед гималаями библиотек в положении золотоискателя, которому надо отыскать крупинки золота в массе песка. В связи с этим стала развиваться теория информационного поиска, нашедшая практическое применение в создании информационно-поисковых систем (ИПС) и устройств. Различают ручные, механизированные и автоматизированные информационно-поисковые си-

стемы. Первые информационно-поисковые устройства представляли собой технические средства, обеспечивающие отбор нужных документов путем механического сопоставления поисковых образов документов с запросами, где применялись перфокарты, суперпозиционные карты, перфокарты с краевой перфорацией и т. п. Внедрение в информационный поиск вычислительной техники породило автоматизированные ИПС (АИПС). Развитие средств телекоммуникаций привело к появлению распределенных информационно-поисковых систем и созданию единой мировой информационной инфраструктуры. В настоящем разделе дается описание того класса автоматизированных информационных систем, которые предназначены для поиска текстов (документов, их частей, фактографических записей) по формальным запросам. Их отличие от других автоматизированных систем заключается в том, что они не выполняют никакой завершающей обработки, связанной с осмыслением полученных результатов или с принятием решений. ИПС — это совокупность средств и методов, предназначенных для хранения и поиска текстов (документов) или данных (фактов).

Сущность информационного поиска. Основные термины. Сущность информационного поиска — это отождествление содержания документа и запроса. Сравнить содержание двух текстов значит определить степень их смысловой близости. Центральным понятием информационного поиска (наряду с понятиями документа и запроса) является понятие релевантности (от англ. relevant — подходящий, имеющий отношение к делу). Под релевантностью обычно понимают степень соответствия одного текста другому, в данном случае документа — запросу. При этом различают смысловую релевантность — действительное соответствие документа, выданного системой, информационнному запросу — и формальную релевантность — соответствие поискового образа документа поисковому предписанию. Под поисковым образом документа и поисковым предписанием здесь понимается представление основного содержания документа и запроса на специальном информационно-поисковом языке (ИПЯ). Наряду со смысловой и формальной релевантностью различают пертинентность (от англ. pertinent — подходящий, относящийся к делу) — соответствие содержания документа фактической информационной потребности. Под информационной потребностью здесь понимается потребность в знании, необходимом конкретному индивиду или коллективу для достижения конкретной цели. Информационная потребность выражается в виде текста на естественном языке, называемого информационным запросом. Информационный запрос не всегда точно выражает информационную потребность. Неумение или неспособность людей сразу адекватно формулиро-

вать свою информационную потребность были отмечены еще Платоном в диалоге «Менон», где Сократ говорит: «Я... вообще знать не знаю, что же такое добродетель. А если я этого не знаю, то откуда мне знать, как этого достичь?». Трудности формулирования информационной потребности объясняются как ее сущностно-субъективным характером, заключающимся в сложности осознания потребителем предмета и степени своего незнания, так и неоднозначностью естественного языка.

На понятии релевантности основываются главные показатели эффективности информационного поиска: полнота и точность. Полнотой поиска называется мера, вычисляемая как отношение количества выданных релевантных документов к общему числу релевантных документов, содержащихся в информационном массиве. Точность поиска — это отношение количества выданных релевантных документов к общему числу документов в выдаче.

В зависимости от объекта и цели поиска различают два вида информационного поиска: документальный и фактографический — и соответственно два типа ИПС — документальные и фактографические. При документальном поиске объектами, выдаваемыми пользователю, являются либо сами документы, содержащие нужные ему сведения, либо их библиографические описания. При фактографическом поиске в качестве объектов хранения и выдачи выступают непосредственно фактические данные (научные, технические, экономические свойства объектов, процессов, явлений и т. п.). Между документальными и фактографическими системами нет непреодолимой разницы. Имеются примеры смешанных систем, в которых документальная и фактографическая информация относится к одному объекту, и фактографическая информация используется для получения дополнительных возможностей при поиске документов. Разновидностью документального поиска с элементами фактографии можно считать библиографический поиск, осуществляемый по элементам библиографического описания документов (год, место издания, вид издания, автор и т. п.). Документы, хранимые в документальных ИПС, и представляют собой, как правило, библиографические описания с рефератами. Запрос к таким массивам может носить вполне фактографический характер, например: выдать сведения о книгах (признак «вид издания» = монографический), такого-то автора (признак «автор» равняется фамилии автора) за последние 5 лет (признак «год издания» больше, чем 1991).

Поскольку в основе информационного поиска лежит сопоставление информации, содержащейся в документах и запросах, возникает проблема представления содержания документов и запросов в поисковой системе.

Для этого создаются ИПЯ — специализированные языки, предназначенные для описания главных тем и формальных характеристик документов и запросов с целью последующего их сопоставления. Все ИПЯ, применяемые в современных АИПС, условно можно разделить на три типа:

1) ИПЯ классификационного типа, в основе которых лежит систематическая классификация понятий;

2) ИПЯ дескрипторного типа, в основе которых лежит перечень лексических единиц, выражаемых словами или словосочетаниями;

3) ИПЯ объектно-признакового типа, в основе которых лежат модели знаний об объектах в конкретных предметных областях, описанные в терминах «признак — значение».

Процесс представления содержания документа или запроса на ИПЯ называется индексированием. При использовании ИПЯ для выражения содержания документа происходит его обобщение, «сжатие». Запись основного содержания документа на ИПЯ называется поисковым образом документа (ПОД). Зафиксированная на естественном языке информационная потребность носит название информационного запроса, а представление запроса на ИПЯ называется поисковым образом запроса (ПОЗ), или поисковым предписанием (ПП).

Информационный поиск, таким образом, сводится к выявлению в поисковом массиве ПОД, соответствующих ПП. Степень соответствия определяется специальными формализованными правилами, называемыми критерием выдачи, или критерием смыслового соответствия (не путать со смысловой релевантностью — речь идет о формальных правилах и формальной релевантности).

Простейшим критерием выдачи является полное совпадение ПОД и ПОЗ, или, точнее, полное вхождение ПОЗ в ПОД, однако на практике, как правило, используются критерии, основанные на частичном совпадении ПОД и ПОЗ.

Часто требования критерия выдачи фиксируются непосредственно в ПП, и поэтому можно говорить о двух версиях одного и того же ИПЯ — ИПЯ для документов и ИПЯ для запросов. В этом случае язык запросов включает в себя дополнительные специальные средства, относящиеся к процедурам поиска и критерию выдачи.

ИПС — это всегда человеко-машинные системы, составной частью которых выступают люди, участвующие в различных технологических процессах (пользователи, обслуживающий персонал). Поэтому различают абстрактную ИПС (собственно логико-семантический аппарат) и полную ИПС (или информационно-поисковую службу). Последняя включает в себя и получателя информации, обычно называемого пользовате-

лем, и способы информационной коммуникации между службой и пользователем. Информационная коммуникация может быть реализована путем выдачи всей релевантной информации, имеющейся в системе, или только новой, текущей, введенной в систему за какой-то последний промежуток времени. Соответственно различают режимы информационного обслуживания: режим «запрос — ответ» и режим избирательного распространения информации (ИРИ). Первый из них предусматривает разовые запросы, поиск по которым проводится, как правило, во всем ретроспективном массиве документов, поэтому этот режим обычно называют «ретро». Поиск в режиме ИРИ — это поиск в массиве новых поступлений документов по постоянным (действующим в течение определенного срока) запросам; их еще называют «профилями».

С точки зрения общения человека и ЭВМ различают пакетный и диалоговый режимы поиска. Первый более характерен для ИРИ, второй — для «ретро». Поиск в пакетном режиме выполняется автономно, без участия человека, как правило, одновременно по всем запросам. Диалоговый поиск предполагает, что в процессе поиска фиксируются те или иные промежуточные результаты и по ходу решения задачи человек (пользователь или специально обученный специалист) ведет с ЭВМ диалог — т. е. принимает некоторые решения, уточняющие запрос или стратегию поиска. Наконец, полученные так или иначе результаты поиска — документы, выдаваемые пользователю, могут представлять собой либо вторичный документ (библиографическое описание, возможно, с рефератом), либо первичный, исходный документ. В этом случае говорят об одноконтурной или двухконтурной системе. Под первым контуром чаще всего имеют в виду хранящиеся в машине поисковые образы документов. Сами документы — второй контур — обычно хранятся вне машины, в привычном бумажном виде или на микроносителях (микрофильмы, микрофиши). В последнее время все большее распространение получают полнотекстовые ИПС, когда и поисковый и документальный контуры содержатся непосредственно в ЭВМ, часто на оптических дисках.

Приведем общую схему классификации ИПС по различным основаниям (см. рисунок на с. 339).

ИПС делят на составные части (подсистемы) по функциональному признаку: ввод документов, ввод и корректировка запросов, поиск, ведение словарей, ведение статистики, обработка результатов поиска и выдача документов.

Различные средства и методы, выполняющие указанные функции, получили название обеспечивающих подсистем, или «обеспечения». Как правило, выделяют следующие основные виды обеспечивающих подсистем: информа-

ционное обеспечение; лингвистическое; техническое; программное; технологическое.

Информационное обеспечение — это информационные массивы документов и запросов, а также средства и способы их описания, построения и классификации.

Лингвистическое обеспечение — это логико-семантический аппарат, состоящий из ИПЯ, правил индексирования и критерия выдачи.

Техническое обеспечение — это технические средства, обеспечивающие хранение и поиск информации.

Программное обеспечение — это программные средства, реализующие все функции ИПС, выполняемые с помощью ЭВМ.

Технологическое обеспечение — это описание автоматизированных и неавтоматизированных процессов и процедур обработки информации в ИПС.

Информационно-поисковый язык. Лингвистическое обеспечение (ЛО) ИПС в целом — это набор логических и лингвистических средств и методов по обеспечению основной задачи информационного поиска. ЛО включает в себя прежде всего языковые средства — специальные языки (в том числе ограниченный естественный язык), правила перевода на эти языки и методы их использования. Для реализации ИПЯ разрабатываются различные словари, форматы и лингвистические алгоритмы. В понятие ЛО входят также критерий выдачи и средства аппарата математической логики и теории множеств, используемые в ИПС.

Основой ЛО является ИПЯ. В отличие от ряда других знаковых систем, выполняющих познавательную и коммуникативную функции в процессе человеческой деятельности, ИПЯ изначально ориентирован на формализованный информационный поиск, и в этом смысле можно говорить об особой информационно-поисковой функции, не связанной с пониманием.

К идеальному ИПЯ предъявляются следующие основные требования: 1) достаточная семантическая сила (способность с необходимой полнотой и точностью описывать основную тему (темы) документов и запросов); 2) однозначность (каждая запись на ИПЯ имеет один смысл, и каждое понятие получает единообразное представление средствами ИПЯ); 3) открытость (возможность пополнения и корректировки лексических средств языка).

Любой письменный язык состоит, по крайней мере, из трех компонентов: алфавит, словарь и грамматика. Это применимо и к ИПЯ. Система графических знаков, используемая для записи лексических единиц (ЛЕ) различных ИПЯ, — алфавит — может включать в себя символы латинского и русского алфавитов, арабские цифры, знаки пунктуации и специальные символы. Совокупность всех ЛЕ ИПЯ называется словарем. В качестве ЛЕ выступают понятия или классы понятий.

Грамматика ИПЯ представляет собой совокупность правил выражения отношений между ЛЕ. Эти отношения бывают двух видов: синтагматические (текстуальные) и парадигматические (базовые). Синтагматические отношения — это отношения между ЛЕ, возникающие в определенной ситуации, в определенном контексте (т. е. в данном документе или запросе).

Парадигматические отношения обусловлены наличием логических связей между понятиями как элементами, отражающими объекты и явления реального мира. Собственно говоря, грамматике в узком, естественноязыковом смысле принадлежат отношения только первого типа. ИПЯ, в которых есть средства для выражения синтагматических отношений, называют «ИПЯ с грамматикой», а те, в которых таких средств нет, — «ИПЯ без грамматики». Парадигматические отношения отно-

сятся к словарному составу языка и, в отличие от естественных языков, в словаре отображаются в явном виде.

Дадим общую классификацию ИПЯ, основанную на понятии координации. Как уже говорилось, в качестве ЛЕ ИПЯ выступают понятия, а точнее, классы понятий. Операция построения новых сложных классов путем пересечения (логического умножения) более простых классов понятий в теории ИПЯ называется *координацией*.

Если ИПЯ имеет такую структуру и правила индексирования, при которых индексирование документов производится приписыванием им предварительно построенных сложных классов, а имена этих классов становятся поисковыми образами, то такой ИПЯ называется предкоординированным. Если же поисковый образ документа состоит из простых классов, а операция координации (построение сложных классов путем пересечения простых) реально происходит только при поиске (после индексирования), то такой ИПЯ называется посткоординируемым. С учетом высказанных положений общая классификация ИПЯ имеет следующий вид:

1. Предкоординированные ИПЯ.

1.1. Иерархические классификации.

1.2. Алфавитно-предметные классификации.

1.3. Фасетные классификации.

2. Посткоординируемые ИПЯ.

2.1. Дескрипторные языки.

2.2. Семантические языки.

2.3. Синтагматические языки.

В предкоординированных ИПЯ заранее перечислены имена всех (или всех основных) классов. Для каждого имени дается название понятия (на естественном языке), которому это имя соответствует. Наиболее известны и распространены *иерархические классификации*, в числе которых в качестве примера можно привести универсальную десятичную классификацию (УДК), Рубрикатор ГАСНТИ, Международную классификацию изобретений (МКИ) и др. В основе построения таких классификаций лежит деление понятий, а сами классификации имеют структуру «дерева знаний» с поименованными узлами. В классификациях каждый класс обозначается кодом и имеет постоянное, определенное место относительно других классов. Основные недостатки иерархических классификаций как ИПЯ заключаются в следующем: 1) невозможность обеспечить многоаспектное индексирование; 2) недостаточная глубина деления; 3) постоянное отставание от развития науки и техники. Иерархические классификации в основном используются для систематизации литературы в библиотеках и книгохранилищах.

Алфавитно-предметные и фасетные классификации также основываются на предварительном составлении перечня основных категорий и классов предметов, встречающихся в конкретной предметной области.

В алфавитно-предметных классификациях классы понятий называются словами естественного языка и располагаются в алфавитном порядке. Алфавитно-предметные классификации предназначены для узкопредметного поиска документов, главная тема которых обозначается соответствующим предметным заголовком, и применяются в основном для составления указателей к систематическим каталогам.

В основу фасетных классификаций положен принцип классификации одного и того же множества предметов и явлений по разным основаниям, а именно: для одного и того же множества объектов строится несколько иерархических деревьев. Фасетные классификации уже позволяют производить многоаспектное индексирование. Основные категории — основания для классификации — называются фасетами (от англ. facet — аспект). Так, множество терминов, обозначающих различные материалы, образуют фасет «Материалы», множество терминов, обозначающих процессы, — фасет «Процессы» и т. д. Каждый термин фасета называют фокусом и для удобства обозначают каким-либо кодом (шифром). Множество фасетов называют схемой классификации.

Например, для классификации кинофильмов можно выделить 4 фасета: жанр, метраж, цвет и формат. В каждом из фасетов перечисляются термины, характеризующие фильмы с точки зрения данного аспекта. Тогда поисковый образ любого фильма будет состоять из четырех терминов-признаков, выбранных из четырех фасетов. Расположение фасетов в схеме классификации и соответственно в ПОД фиксированно. Эта фиксированная последовательность называется фасетной формулой. В то же время в связи с большой трудоемкостью разработки фасетных классификаций они могут быть созданы только для узких предметных областей. Методика индексирования с применением фасетных классификаций также сложна. Поэтому на практике в автоматизированных ИПС эти классификации широкого применения не нашли.

Также не нашли практического применения семантические и синтагматические языки. Наиболее известны среди них семантический код Перри — Кента, язык *RX*-кодов и язык СИНТОЛ. Их разработчики основную задачу видели в однозначном переводе с естественного языка на ИПЯ с учетом семантических и синтаксических факторов. Характерная особенность этих языков — наличие средств, явно описывающих семантическую и семантико-синтаксическую структуру понятий и выражений (семантические множители, термы, реляторы, предикаты, сущности и т. п.).

Дескрипторные языки. Развитие автоматизированных ИПС в основном оказалось связано с ИПЯ дескрипторного типа.

В основу построения дескрипторных языков положен принцип координатного индексирования, заключающийся в том, что главная тема документа описывается перечнем дескрипторов (англ. от describe — описывать) — слов, являющихся именами простых понятий. Такие слова выступают в качестве как бы координат документов в некотором умозрительном *n*-мерном предметно-тематическом пространстве. Принцип координатного индексирования был разработан в 50-х годах в процессе создания механизированных ИПС. Одной из первых таких ИПС была система УНИТЕРМ, разработанная в США М. Таубе. В ней в качестве индексов, описывающих содержание документов и запросов, использовались ключевые слова, выбранные из текста. Все такие слова (унитермы) имели одинаковый иерархический ранг. Словарный состав унитермных языков не разрабатывался предварительно, а формировался непосредственно в процессе индексирования. Координация, или логическое умножение понятий, в результате которой из простых лексических единиц строятся более сложные, выражающие более узкие понятия, осуществлялись как бы дважды: потенциально (мысленно) — при индексировании документа и реально — в процессе поиска: при сопоставлении поискового образа документа с терминами поискового предписания.

Например, пересечением понятий *библиотеки* и *автоматизация* порождается новое более узкое понятие *автоматизация библиотек*. Логическое умножение понятий хорошо иллюстрируется на кругах Эйлера.

При индексировании документов и запросов ключевыми словами, т. е. терминами, взятыми непосредственно из индексируемых текстов, возникла потребность в лексическом контроле (ЛК) за словарным составом ИПЯ. ЛК — это совокупность мер, принимаемых при координатном индексировании с целью сокращения до минимума отрицательных последствий неоднозначного и неединообразного употребления слов. Основная цель ЛК — обеспечить однозначное представление понятий в ПОД и ПОЗ, вне зависимости от способа выражения их в текстах. Термин *лексический контроль* имеет синоним *словарный контроль*, и это говорит о том, что для нормирования ЛЕ ИПЯ используется словарь индексирования, или дескрипторный словарь, в котором в явном виде перечислены все важнейшие ключевые слова и дескрипторы. Дескрипторный словарь с зафиксированными в нем парадигматическими отношениями называется информационно-поисковым тезаурусом — ИПТ (от греч. thesauros — запас).

Суть ЛК, осуществляемого с помощью тезауруса, в основном сводится к следующему: 1) контроль синонимии; 2) разрешение омонимии; 3) нормализация (сведение воедино) всех вариантов одного и того же слова.

Кратко коснемся каждой из составляющих ЛК.

1. Контроль синонимии является, может быть, главной функцией ЛК в ИПС. Он заключается в том, что все синонимы заменяются одним и тем же лексическим вариантом. Этот вариант называется дескриптором и выступает в качестве ЛЕ дескрипторного языка. Совокупность дескриптора и близких ему по смыслу слов образует класс условной эквивалентности. В число синонимов включаются:

— полные синонимы (*радар — радиолокатор, ЭВМ — компьютер*);

— фонетические (графические) варианты слов (*секстан — секстант*);

— аббревиатуры (*ИПС — информационно-поисковая система*);

— слова, совпадающие по смыслу в одном из значений (*компьютер — машина, статья — работа*).

Иногда к синонимам относят и антонимы («квазисинонимы») (*жесткость — мягкость, фокусировка— дефокусировка*).

2. ЛК также призван различать омонимы (омографы) и многозначные слова. Это достигается присвоением им различных кодов (в ИПЯ, лексика которых записывается с помощью цифрового алфавита) или системой помет, например: *конденсаторы* (для пара) — *конденсаторы* (электрические), *литье* (*предмет*) — *литье* (процесс), *меркурий* (металл) — *Меркурий* (планета). В ряде исследований, однако, показано, что неразличение омонимов не ведет к заметному информационному шуму.

3. В процессе индексирования все ключевые слова, представленные в текстах словоформами, заменяются на нормализованную словарную форму этого ключевого слова (дескриптора) (например, им. падеж ед. числа). Иногда в один класс условной эквивалентности сводятся не только грамматические формы одной лексемы, но и разноосновные и даже относящиеся к разным частям речи лексемы, объединяемые общим лексическим значением («гиперлексемы»), например: *трелевка — трелевочный, сверло — сверление.*

После лексического контроля вторая важнейшая функция тезауруса — повышение эффективности ИПС, т. е. полноты и точности. Это достигается использованием парадигматических и синтагматических отношений.

Для фиксации синтагматических отношений, отражающих связи между ЛЕ в тексте, вводятся специальные грамматические средства — указатели роли и связи. Указа-

тели роли определяют, какое значение имеет в тексте (и соответственно в ПОД) тот или иной дескриптор, и играют роль грамматических категорий. В качестве примеров таких категорий можно назвать «процесс», «свойства», «материя», «оборудование», «среда». Обозначаются указатели посредством соответствующих кодов, приписываемых дескрипторам в ПОД. Указатели связи определяют, какие из дескрипторов в ПОД связаны между собой (логически или синтаксически). Обозначаются они также особыми символами, приписываемыми дескрипторам в ПОД.

Использование указателей роли и связи позволяет повысить показатель точности. Однако при этом снижается полнота поиска. Индексирование документов с применением указателей роли и связи требует дополнительных трудозатрат и влечет за собой возможность ошибок. Все это привело к тому, что на практике указатели роли и связи, как правило, перестали использоваться. Их отсутствие компенсируется другими способами.

Парадигматические (базисные) отношения отражают смысловые связи между ЛЕ ИПЯ, зависящие не от контекста, а от отношений между объектами в реальном мире. К их числу относятся иерархическое отношение (род — вид), а также многочисленные отношения, получившие название ассоциативных (часть — целое, предмет — свойство, процесс — результат и др.). Смысловые (логические) связи между понятиями в явном виде (эксплицитно) зафиксированы в тезаурусах. ИПТ — это семантическая модель плана содержания соответствующей области знания.

Разработка ИПТ представляет собой большую и сложную работу. К основным этапам создания ИПТ относятся:

— определение тематического охвата ИПТ;

— сбор текстового материала;

— отбор ключевых лексических единиц (формирование словника);

— группирование ЛЕ в классы условной эквивалентности;

— установление парадигматических отношений между ЛЕ, выбор их типов;

— построение словарных статей и указателей;

— оформление ИПТ;

— экспертиза и регистрация.

При составлении ИПТ в качестве вспомогательного инструмента, как правило, используется ЭВМ. Правила разработки, структура и форма представления ИПТ определяется государственными стандартами (ГОСТ). При оформлении тезауруса исходят из требования удобства пользования. Форма представления его должна обеспечивать нахождение нужной

по смыслу ЛЕ и всех ЛЕ, с ней связанных. Чаще всего ИПТ имеет две части.

1. Словарная часть (собственно тезаурус), представляющая собой алфавитный список дескрипторов вместе с их словарными статьями (гнездами). Словарная статья помимо самого дескриптора обычно содержит четыре подмножества:

— ключевые слова-синонимы;

— «вышестоящие» дескрипторы (находящиеся с данным в отношении «род — вид»);

— «нижестоящие» дескрипторы (находящиеся с данным в отношении «вид — род»);

— ассоциативные (ассоциированные) дескрипторы.

2. Указатели различных видов, облегчающие пользование тезаурусом (алфавитный указатель всех ключевых слов, пермутационный указатель для элементов словосочетаний, частотный указатель и др.).

Следует заметить, что частной, но одной из труднейших проблем при составлении тезауруса является задача выбора словосочетаний для включения в ИПТ.

Процесс индексирования, т. е. выбор дескрипторов для включения в ПОД и ПОЗ, осуществляемый вручную, определяется специальными инструкциями, называемыми методикой индексирования. Тем не менее не будет большим преувеличением сказать, что индексирование — это искусство. Эффективность поиска во многом определяется качеством индексирования, которое зависит как от качества ИПТ, так и от методики индексирования. В оценке качества индексирования, как и при поиске, используют показатели полноты (глубины) и точности (детальности) индексирования. Полнота определяется, в первую очередь, количеством дескрипторов, включенных в ПОД (ПОЗ), точность — так сказать, их «качеством», т. е. смысловой близостью выбранных дескрипторов к основному содержанию документа (запроса). Эти характеристики находятся в отношении обратной зависимости. Для обеспечения полноты индексирования и поиска применяется *избыточное индексирование*. Под избыточным индексированием понимается дополнение ПОД и ПОЗ дескрипторами, связанными по смыслу с основными.

Проблема автоматизации индексирования приобрела особую актуальность в связи с созданием документальных баз данных большого объема, необходимостью обмена информацией и задачей повышения эффективности поиска. В условиях дефицита ресурсов для ручного индексирования все возрастающего входного потока документов требуется все большее число квалифицированных специалистов, при том, что качество «ручных» ПОД оказывается весьма невысоким. Начало ис-

следованиям по автоматизации индексирования положили работы Г. Луна (H. Luhn) и П. Баксендейла (P. Baxendale), а в нашей стране — Д. Г. Лахути.

В настоящее время созданы системы автоматического индексирования (АИ), не уступающие по качеству обеспечиваемого ими поиска аналогичным системам с ручным (интеллектуальным) индексированием. В общем случае на стадии АИ происходит отбор ЛЕ из входных документов для включения в ПОД и нормализации этих единиц. Отбор производится по положительному (словник значимых слов) или отрицательному словарю (перечень незначащих слов, «стоп»-слов). Нормализация ЛЕ при АИ производится на основе методов морфологического анализа путем приведения ЛЕ к каноническому виду. Также имеются способы нормализации словоформ непосредственно в процессе поиска. Наиболее распространен механизм, называемый усечением (или маскированием), когда в запросе в качестве поискового термина указывается только значимая часть слова (основа или квазиоснова).

Трудности, связанные с решением проблемы качественного индексирования, ускорили появление ИПС, обеспечивающих поиск информации непосредственно, по текстам рефератов и библиографическим описаниям. Одной из заметных тенденций в информационном поиске является широкое распространение систем поиска на основе ограниченного естественного языка. Их называют также системами бестезаурусного типа. Класс бестезаурусных ИПС в основном отличают две особенности:

— отказ от лексического контроля;

— автоматизация индексирования документов.

Для координатного индексирования в таких ИПС используются слова и словосочетания естественного языка. Фактически мы имеем дело с дескрипторными ИПЯ без фиксированной парадигматики и без фиксированного словаря. А. И. Черный считает, что в этом случае используются два варианта ИПЯ: один — с нулевой нормализацией ключевых слов и словосочетаний естественного языка (для координатного индексирования документов), другой — с большой степенью логической обработки (для избыточного индексирования запросов) (Черный, 1975).

Указанные системы берут свое начало от системы УНИТЕРМ. О функциях и преимуществах использования тезаурусов уже говорилось. Однако и системы без контроля лексики имеют свои достоинства. Прежде всего, это их высокая точность. Очевидно, что никакой тезаурус не в состоянии охватить все многообразие лексики естественного языка (ЕЯ), поэтому можно утверждать, что процесс индексирования ведет к обеднению, а в ряде случаев и к искажению содержания до-

кументов. В большинстве систем, как правило, преобладают узкие многоаспектные запросы. Системы на базе ЕЯ в этом случае имеют огромное преимущество: они обладают полной специфичностью. Тезаурус никогда не может быть настолько специфичен, как ЕЯ. Кроме того, разные ИПТ, как правило, несовместимы. Проблема совместимости тезаурусов одно время представлялась одной из самых главных проблем в ИПС. Предлагались различные подходы, в том числе и создание промежуточного языка. В целом она так и осталась нерешенной. Развитие систем на базе ЕЯ попросту снимает эту проблему.

На наш взгляд, важнейшим событием, стимулировавшим процесс развития указанных систем, явился известный второй крэнфилдский эксперимент (Великобритания, 1966 г.). Это и другие исследования, в том числе и в нашей стране, показали, что ряд поисковых механизмов (фиксация нетривиальных парадигматических и синтагматических отношений, учет синтаксических связей и т. п.) или не приводят к существенному улучшению показателей поиска, или, незначительно улучшая одну из основных характеристик (полнота, точность), заметно ухудшают другую. Поэтому уместно говорить о замене одних поисковых средств другими, более простыми и дешевыми. Целью крэнфилдского эксперимента как раз и была сравнительная оценка эффективности различных средств и стратегий поиска.

Эксперименты по оценке систем показали, что системы поиска по тексту, будучи намного более экономичными, часто по показателям не уступают системам тезаурусного типа. В основном наблюдается повышение точности и снижение полноты. И это закономерно. Для того чтобы добиться лучших показателей при поиске с использованием ЕЯ, необходимы меры, компенсирующие известные недостатки ЕЯ. ИПС можно рассматривать как совокупность двух механизмов: смыслоразличения и смыслоотождествления. Уровень «мощности» этих механизмов должен соответствовать друг другу, и как только мы увеличиваем смыслоразличительную силу ИПС (ИПЯ), мы должны обязательно усилить механизм смыслоотождествления. Лексический контроль, нормируя лексику, облегчает выполнение функции смыслоотождествления. И естественно, отказ от него влечет за собой снижение полноты поиска при увеличении точности.

Ввод в ИПС полных текстов документов (рефератов) позволяет полностью сохранить все лексические единицы документа и контекстуальные отношения между ними. Отличие систем поиска по связному тексту от традиционных дескрипторных заключается в том, что в них для каждой словоформы имеется возможность указать ее точное местонахождение в тексте (номер абзаца, номер предложения, номер слова в

предложении). Наличие точных «адресов» словоформ породило специальные контекстные операторы языка запросов и алгоритмы поиска, которые используют расстояние между словоформами и другие позиционные характеристики («позиционные операторы»). Это позволяет использовать синтагматические отношения ЛЕ (ср.: «указатели связи»), но более простым и гибким способом. ИПС подобного рода получили название систем с линейной (или позиционно-скобочной) грамматикой. В качестве примеров подобных систем можно назвать пакеты прикладных программ ДИАЛОГ, CDS/ISIS, СКОБКИ, РАСПРИ. Использование позиционных операторов в сочетании с режимом диалогового поиска позволяет достаточно простыми средствами избавляться от ложной координации терминов, неизбежной при поиске по тексту значительного объема.

Проблема смыслоотождествления (и повышения полноты) в бестезаурусных ИПС решается в основном на стадии составления ПП, где осуществляется избыточное индексирование запросов. Там каждому термину из ПП приписываются синонимичные или близкие ему по смыслу термины, в том числе термины, находящиеся с данным в родо-видовых или других парадигматических отношениях, т. е. происходит как бы «развертывание» словарной статьи ИПТ. Отличие от индексирования в традиционном смысле заключается в том, что исходные ЛЕ запроса (и тем более документа) здесь никак не «портятся» и что методика индексирования запроса при таком подходе в большей степени способна учесть конкретную информационную потребность. В ИПС бестезаурусного типа тезаурус фактически не исключается, но он начинает использоваться, как и тезаурус в лингвистике, в качестве средства для моделирования информационной потребности и смыслового варьирования запросов. Важным направлением в развитии документальных ИПС является разработка методов и создание подсистем автоматизированного формирования качественных ПП и их автоматической корректировки на основе данных обратной связи по релевантности.

Важнейшей составной частью ЛО ИПС являются разного рода словари. Они используются для индексирования, в том числе автоматического, для автоматизированного обнаружения ошибок при вводе документов и запросов, при поиске и в других случаях.

Методы автоматизированного обнаружения ошибок в текстах можно разделить на две большие группы: методы полиграммного контроля и методы словарного контроля. В первом случае в качестве средства обнаружения ошибок используются списки допустимых в данном языке двух- или трехбуквенных сочетаний. Во втором случае слова текста подвергаются морфологическому анализу и сравниваются с эталон-

ным словарем. Как показала практика, словарные методы работают намного эффективнее. Но их эффективность зависит от объема словаря и полноты покрытия анализируемых текстов. Центральное место в подсистемах обнаружения ошибок занимают процедуры автоматизированного морфологического анализа, учитывающие явления словоизменения и словообразования.

Отдельная важная часть ЛО ИПС — средства и методы автоматизации составления и ведения машинных словарей. Обычно задачи этого рода объединяются в понятие автоматизированной словарной службы.

Критерий выдачи. При сопоставлении документов и запросов требуется определить релевантность документа по отношению к запросу и принять решение о выдаче или невыдаче документа на данный запрос. Это решение принимается в соответствии с критерием выдачи (КВ). Критерий выдачи (называемый также критерием смыслового соответствия — КСС) — это формальные правила, на основе которых определяется степень формальной релевантности документа и запроса, т. е. соответствие ПОД и ПОЗ. В дескрипторных ИПС все применяемые КВ можно разбить на две большие группы: вычисляемые и логические.

Вычисляемые КВ оперируют количественными параметрами, в том числе объемными характеристиками ПОД и ПОЗ. Так, ряд критериев основывается на теоретико-множественной операции пересечения множеств терминов ПОД и ПОЗ. Величина такого пересечения может определяться в процентах. Простейший случай — полное совпадение ПОД и ПОЗ. Однако это требование является слишком жестким. На практике обычно задается какое-то пороговое значение (например, 50%), выше которого находится уровень релевантности. Простые критерии выдачи исходят из равнозначности терминов, входящих в ПОД и ПОЗ. В этом случае величина значения критерия выдачи определяется по формуле

$$R = \frac{N_i}{N_z} \cdot 100\%,$$

где N_i — количество совпавших при поиске терминов ПОД и ПОЗ; N_z — общее количество терминов в ПОЗ.

В то же время часто разные термины, особенно в ПП, не равнозначны. В связи с этим появилось понятие «весового» коэффициента, или «веса» термина. Операцию присвоения терминам весов стали называть «взвешиванием» терминов. Способы взвешивания могут быть разные. Самый простой из них заключается в том, что пользователь при составлении ПП оценивает каждый дескриптор ПП определенным числом — весовым коэффициентом. Кроме того, пользователь может задать и пороговое значение суммы весов. Также имеются автоматизированные методы взвешивания терминов, в том числе при

автоматическом индексировании документов. Так, например, может анализироваться относительная частота встречаемости какого-либо слова. Очевидно, что слова, относительная частота которых в данном документе превосходит относительную частоту для этих слов в совокупности всех документов, являются более важными с точки зрения основного содержания данного документа. Другой подход заключается в приписывании повышенных весов терминам, выбранным для включения в ПОД из наиболее важных зон документа (заглавия, резюме и т. п.).

Собственно КВ, который определяет степень соответствия ПОД и ПОЗ, может быть вычислен самыми разными способами. В общем случае с помощью метода сравнения векторов (множества взвешенных терминов запроса и документа) для каждой пары «документ — запрос» подсчитывается значение некоторого коэффициента подобия. Это дает возможность ранжировать (эшелонировать) документы выдачи в порядке уменьшения коэффициента подобия, а также выдавать то или другое количество документов в зависимости от заданного порогового значения.

Все эксперименты показывают высокую эффективность ИПС с весовым КВ. Однако на практике постепенно повсеместное распространение получили ИПС с логическим КВ, когда ПОЗ строятся с использованием логических (булевых) операторов И, ИЛИ, НЕ (также обозначаются знаками &, $\vee$, $\neg$). Это дает основание говорить о двух ИПЯ в рамках одной ИПС — под вторым имеется в виду специальный язык запросов, базирующийся на булевой алгебре, возможно, с дополнительными позиционными операторами. В этом случае ключевые слова или дескрипторы ПП играют роль булевых переменных, принимающих значение 1 («истина»), если данное слово находится в ПОД, и 0 («ложь»), когда оно там отсутствует. Документ признается релевантным запросу, если логическая формула запроса в целом получает для данного документа значение «истина».

Чаще всего поисковое предписание записывается как конъюнктивная нормальная форма (конъюнкция дизъюнкций). Оператором ИЛИ (дизъюнкция) обычно связывают термины, находящиеся в отношении синонимии. Оператором И (конъюнкция) связывают термины (или группы терминов), выражающие разные аспекты запроса. Использование булевых операторов обеспечивает логику сравнения документов и запросов, понятную пользователю. Поиск (вычисление истинности для элементов ПП), как правило, проводится по инвертированным файлам (наборам данных), построенным на основе словника документального массива, и характеризуется высокой скоростью.

В то же время в ИПС с логическим КВ не обеспечивается

ранжирование выдаваемых документов, трудно обеспечить контроль за объемом выдачи, возникают трудности при обеспечении желаемого уровня эффективности ИПС — и поэтому безоговорочная «победа» логического критерия выдачи представляется немного загадочной и, возможно, не окончательной. Один из пионеров информационного поиска в США Ф. Ланкастер пишет: «Возможно, что использование булевой алгебры для построения автоматизированной поисковой системы было ошибкой... источник которой лежит в сходстве первых поисковых автоматизированных систем с системами, построенными на картах с краевой перфорацией или на принципе оптического совпадения... В этих системах не предусматривалась возможность частичного совпадения и выдачи списков документов, что же касается автоматизированных систем, то в них такого рода ограничения не имеют смысла» (Ланкастер, 1972, с. 267).

Информационное обеспечение. В ИПС различают первый и второй документальные «контуры». Под первым контуром понимаются поисковые массивы (ПОД, ПОЗ), хранящиеся в ЭВМ (на внешних носителях), под вторым — массивы первичных документов, хранящихся обычно вне ЭВМ (на бумаге или на микроносителях). Однако рост емкости запоминающих устройств ЭВМ, появление оптических дисков сверхбольшой емкости, подключаемых к ЭВМ, ведут к тому, что растет число полнотекстовых баз данных, в результате чего первый и второй контуры как бы сливаются.

Современные ИПС ориентируются на широкий обмен информационными массивами (базами данных). В связи с этим их представление на машинных носителях — формат — регламентируется специальными международными и государственными стандартами. В настоящее время наиболее распространен международный стандарт ISO 2709 (ГОСТ 7.14—84). Все документы, записанные в соответствии с указанным стандартом, имеют единую структуру:

— маркер записи — поле фиксированной длины в начале записи, содержащее параметры, относящиеся ко всей записи;

— справочник — совокупность элементов оглавления для каждого библиографического элемента данных, представленного в документе;

— область данных, содержащая элементы данных.

Наполнение каждого документа (имена библиографических элементов данных, их семантика, форма представления) также подлежит унификации. Здесь наибольшее распространение получили семейства MARC (USMARC, UNIMARC и др.).

Техническое и программное обеспечение. Техническое и программное обеспечения представляют собой средства реализации поисковой функции и будут рассмотрены здесь только в общих чертах. Также совсем не рассматриваются вопросы технологи

ческого обеспечения, представляющие собой описание технологии эксплуатации конкретных информационно-поисковых служб.

Первые ИПС были ручными и механизированными. Все современные ИПС реализуются на базе ЭВМ. Основные требования ИПС к комплексам технических средств — это быстродействие и емкость запоминающих устройств. Возможности электронной техники в последние годы выросли чрезвычайно. Быстродействие современных ЭВМ— миллионы операций в секунду — и память на магнитных и оптических дисках, измеряемая мега-, гига- и терабайтами, обеспечивают создание практически любых поисковых систем. Долгое время узкое место в ИПС представляли процедуры ввода и вывода информации. В последние годы достигнуты большие успехи в области бесперфорационного ввода с помощью фотосчитывающих устройств (сканеров) в совокупности со специальными программами распознавания шрифтов. Для выдачи информации используются видеотерминальные устройства и различного рода принтеры, в том числе лазерные, обеспечивающие быструю печать с высокой разрешающей способностью. В современных ИПС, как правило, обеспечен удаленный поиск с использованием каналов связи на основе сопряжения ИПС со специальными программно-аппаратными комплексами (процессоры телеобработки, модемы, протоколы). В зависимости от объема информации и назначения ИПС создаются как на базе больших универсальных ЭВМ, так и на базе персональных компьютеров, возможности которых сегодня также очень велики. В случае необходимости ПЭВМ объединяются в локальные сети, позволяющие создавать распределенные базы и банки данных.

Программное обеспечение (ПО), поддерживающее функционирование ИПС, условно можно разделить на две части: общесистемное ПО и прикладное ПО. Прикладное ПО реализует все функции конкретных ИПС: ввод информации, контроль, хранение, индексирование, поиск, вывод. Состав программных средств и методы их реализации во многом зависят от лингвистического и информационного обеспечения.

Важное место в разработке ПО ИПС занимает вопрос модели данных и организации информационной базы (поисковых массивов). Выбор способа организации данных зависит от разных факторов, таких, как объем данных, частота использования, характер и динамика изменений, возможности технических средств, режим поиска (ИРИ, «ретро»), логика поиска и критерий выдачи. Среди всего разнообразия способов организации данных следует различать последовательный и прямой методы доступа к данным. Суть последовательного доступа заключается в том, что записи подвергаются обработке только последовательно, одна за другой. При прямом методе доступа разрабатываются специальные алгоритмы и, как правило,

строятся вспомогательные файлы (наборы данных), обеспечивающие быстрый доступ к любой произвольно выбранной записи.

Для разработки ПО ИПС могут использоваться алгоритмические языки и средства, предоставляемые операционными системами. На практике в основном используются стандартные программные комплексы специального назначения, а именно, пакеты прикладных программ (ППП) для документального информационного поиска. В качестве примера могут быть названы широко распространенные у нас в стране ППП ДИАЛОГ и CDS/ISIS. Эти пакеты, как правило, обеспечивают возможность построения ядра поисковой системы и включают в себя:

— средства создания БД;

— средства пополнения и корректировки БД;

— средства составления запросов и поиска;

— средства, обеспечивающие диалоговое взаимодействие пользователей с БД;

— средства вывода информации.

Эти пакеты дают возможность строить инвертированные файлы (индексы) для быстрого поиска методом прямого доступа. Инвертированные файлы могут строиться по любым элементам данных, в том числе и по словоформам текстовых полей. Любое слово текста фиксируется в инвертированном файле с указанием его «адреса» в тексте, что позволяет использовать при поиске контекстные (позиционные) операторы, указывающие на необходимость попадания терминов запроса в одну и ту же часть текста (например, предложение). Это особенно важно для полных текстов. Одной из наиболее развитых полнотекстовых ИПС является система АРТЕФАКТ, в которой реализованы морфологическая нормализация словоформ (что крайне необходимо для русского языка), сжатие данных, развитый язык запросов. Контекстные операторы позволяют в запросе строить словосочетания, что существенно повышает качество и гибкость поиска. Для расширения поисковых возможностей имеется механизм поиска по так называемым форматным полям, позволяющий задавать дополнительные условия с помощью арифметических операторов сравнения (больше, меньше, равно и др.).

Механизм поиска по форматным полям с известными ограничениями может быть использован и для фактографического поиска. В то же время для построения фактографических ИПС, а в ряде случаев и документальных, используется другой класс систем — системы управления базами данных (СУБД).

Фактографические информационно-поисковые системы. Ав-

томатизированная фактографическая информационно-поисковая система (АФИПС), как следует из ее названия, — это информационная система, в которой хранится информация о конкретных фактах, свойствах определенных объектов и т. д. При описании АФИПС обычно используют терминологию теории баз данных, так как она достаточно хорошо развита и практические разработки по созданию АФИПС ведутся на основе этой теории. Под базой данных (БД) понимается совокупность хранящихся в памяти ЭВМ данных и связей между ними. В каком-то смысле БД является моделью или информационным отображением «куска реальной действительности». Те явления, факты, предметы и т. п., информация о которых хранится в БД, называют объектами. Иными словами, объект — это элемент, описываемый в БД действительности, который мыслится как единое целое. Так, например, объектами могут быть рейс самолета, деталь механизма, служащий предприятия и т. д. Любой объект обладает различными свойствами, качествами, которые называются признаками или атрибутами. Набор атрибутов определяет объект, а с информационной точки зрения объект — это и есть набор атрибутов. Скажем, объект *рейс самолета* может характеризоваться такими атрибутами, как *номер рейса, тип самолета, пункт прибытия, пункт отправления, дата полета, время отправления, время прибытия, время в полете, цена билета* и т. п.

Если говорить в программистских терминах, то каждый объект представляется в виде записи. Каждая запись состоит из полей, соответствующих атрибутам.

Информационные системы, которые строятся исходя из этой терминологии, иногда называют объектно-признаковыми информационными системами.

Таким образом, БД представляет собой некоторое множество объектов (наборов атрибутов) и связей между ними.

С точки зрения структуры БД обычно делят на три типа: иерархические, сетевые и реляционные.

Иерархические БД — это БД, имеющие структуру дерева. Сетевые БД имеют структуру произвольной сети. Имеются хорошо разработанные математические средства для обработки древесных структур; в то же время работа с сетью представляет много сложностей. Поэтому поиск в иерархической БД производится гораздо эффективнее, чем в сетевой. Однако нельзя удовлетворительно представить произвольную информацию в виде древесной структуры, и поэтому иерархические БД обычно используются только тогда, когда описываемая информация имеет иерархический характер. В последнее время чаще всего используются реляционные БД, которые представляют из себя наборы прямоугольных таблиц. Связи между таблицами обеспечиваются перекрестными ссылками.

Для того чтобы иметь возможность работать с БД, необходимы специальные программы. Комплекс программ, обеспечивающих работу с БД, называется системой управления базами данных (СУБД). Предполагается, что СУБД состоит из достаточно универсальных программ, позволяющих работать с различными БД.

СУБД должна обеспечивать создание структуры БД, модификацию информации, содержащейся в БД, поиск информации в БД и, возможно, ряд дополнительных функций, таких, как, например, выдача информации в удобном для пользователя виде и др. Рассмотрим эти функции СУБД подробнее.

Создание структуры БД. При разработке БД большое внимание должно уделяться разработке ее структуры. Для того чтобы эффективно хранить и обрабатывать информацию, необходимо заранее определить, какая именно информация будет храниться, какие типы объектов должны быть представлены в БД. Необходимо определить для каждого типа объектов тот набор атрибутов, который будет его определять. Необходимо также определить, какие типы связей могут устанавливаться между различными типами объектов. Схема, в которой отражены все типы объектов, указаны для каждого типа объекта его набор атрибутов, а также отображены возможные типы связей, называется общей концептуальной схемой или моделью данных.

Обычно разработка БД начинается с изучения той предметной области, информация о которой должна быть отображена в БД, и с построения общей концептуальной схемы. СУБД должна обеспечивать возможность ввода информации о структуре создаваемой БД на основе общей концептуальной схемы. При этом необходимо определить (в программистском смысле) тип данных каждого атрибута так, чтобы, с одной стороны, любое значение этого атрибута могло быть представлено в виде элемента данных этого типа, а с другой стороны, так, чтобы поле, соответствующее данному атрибуту, не было слишком большим. Например, если одним из атрибутов объекта *служащий* является *фамилия,* то он должен быть представлен в виде строки определенной длины. Подбор длины этой строки — операция достаточно тонкая, так как если длина этой строки будет недостаточна, то длинная фамилия может поместиться не полностью, а если длина будет слишком велика, то каждая запись займет места в памяти несколько больше, чем могла бы, а с учетом того, что этих записей в БД может быть очень много (тысячи), то мы получим неоправданный рост объема требуемой памяти.

Обычно при создании БД считают, что она не должна содержать избыточной информации. Иногда БД даже определяют как неизбыточную совокупность данных. Однако на практике эта неизбыточность данных редко достигается. Часто для

более удобного поиска информации избыточность бывает даже выгодна. Например, если у объекта *рейс самолета* среди атрибутов мы зададим такие, как *время вылета, время прилета* и *время в полете,* то очевидно, что любой из этих трех атрибутов может быть вычислен при помощи двух остальных. Но если мы собираемся осуществлять поиск в БД по каждой из трех характеристик рейса самолета, то для организации эффективного поиска целесообразно оставить все три атрибута. Однако в общем случае следует стремиться сделать избыточность минимальной.

Опасность избыточной информации в БД не только и не столько в том, что она требует дополнительных объемов памяти, а, в первую очередь, в том, что модификация данных требует дополнительного контроля за правильностью изменений. Так, в приведенном выше примере нельзя изменить значение одного из атрибутов, не изменив значения другого. Желательно, чтобы в таких случаях контроль осуществлялся программными средствами; в противном случае достаточно велика вероятность возникновения противоречий в БД, что недопустимо. Поэтому некоторые авторы считают, что в БД может существовать только контролируемая избыточность.

Модификация информации, находящейся в БД. СУБД должна обеспечивать возможность изменения информации, находящейся в БД. Эти изменения могут быть следующих типов: добавление новой записи (нового объекта), устранение существующей записи, изменение значения атрибута в существующей записи. Более сложным изменением является модификация структуры БД. Практически невозможно спроектировать идеальную общую концептуальную схему данных, которая при практической работе не потребует модификации. Часто при использовании БД выясняется, что какие-то типы объектов не нужны, каких-то других типов объектов не хватает, в каких-то типах объектов нужно добавить или, наоборот, устранить какие-то атрибуты, а какие-то связи должны быть установлены иначе. СУБД должна позволять изменять структру БД таким образом, чтобы никакая полезная информация, хранящаяся в БД, не пропадала.

Поиск информации. Одной из главных функций СУБД является обеспечение эффективного поиска информации по запросу пользователя. Предполагается, что поиск будет осуществляться точно и быстро. Для осуществления быстрого поиска необходимо отпределенным образом организовать данные.

Существуют достаточно эффективные способы поиска в упорядоченной последовательности, но если последовательность не упорядочена, то поиск можно осуществлять только полным просмотром всей последовательности, что в случае большого количества записей занимает много времени. Однако упорядочить весь файл записей можно только по значениям одного

атрибута, а поиск необходимо проводить по значениям разных атрибутов. Для обеспечения данной возможности обычно создаются инвертированные индексные файлы, своеобразные указатели, в которых содержатся значения атрибутов и ссылки на нужные записи. Индексные файлы представляют из себя избыточную информацию. Поэтому, для того чтобы эта избыточность была контролируемой, в СУБД должен быть механизм корректировки индексных файлов при модификации информации.

Для каждого атрибута, по значению которого предполагается организация быстрого поиска, должен организовываться свой индексный файл. Таким образом, в СУБД должна быть программа организации индексных файлов. Атрибут, по значению которого возможен быстрый поиск, называется ключом или ключевым атрибутом. Ключ, по которому упорядочен сам файл записей, называется первичным ключом, остальные ключи называются вторичными. Обычно требуют, чтобы первичный ключ был уникальным, т. е. однозначно идентифицировал запись. Значения вторичных ключей в разных записях могут повторяться. Скорость поиска по первичному ключу — максимальная.

СУБД должна обеспечивать поиск не только по значениям ключевых атрибутов, но при поиске информации по неключевым атрибутам работа может осуществляться не столь быстро.

Существует два простых типа запроса на поиск информации в БД. Первый тип — это поиск значения определенного атрибута по известному объекту (первичному ключу). Пример такого запроса: «Найти время вылета самолета, если известен номер рейса (первичный ключ)». Второй тип — это поиск объекта (или объектов) по известному значению определенного атрибута. Пример: «Найти номер рейса (номера рейсов), если известен пункт назначения». Сложные запросы составляются из нескольких простых.

Дополнительные функции СУБД. СУБД обычно обеспечивают выполнение ряда дополнительных функций. Так, чаще всего СУБД может выдавать найденную информацию в удобной для пользователя форме (например, в виде таблицы). Кроме того, СУБД чаще всего обеспечивает выполнение различных вычислительных задач (например, выполнение запроса типа «Сколько объектов в БД удовлетворяют определенным условиям?»).

СУБД должны также выполнять ряд функций, которые необходимы для любых вычислительных систем, такие, как защита информации от системных сбоев, от случайных ошибок пользователя, а, кроме того, при необходимости и от несанкционированного доступа.

Теория баз данных представляет математические и програм-

мные средства для создания АФИПС. В целом же методология создания фактографических систем разработана не так хорошо, как для документального поиска. Отчасти это объясняется широтой спектра АФИПС и большими различиями их между собой. С точки зрения информационного наполнения АФИПС можно сказать, что в ряде применений, связанных с обработкой текстовой информации, фактографические системы смыкаются с экспертными системами и другими задачами автоматической обработки текста.

Литература

Дейт К. Введение в системы без данных. М., 1980.

Информатика / Под ред. К. В. Тараканова. М., 1986.

Ланкастер Ф. Информационно-поисковые системы. М., 1972.

Мартин Дж. Организация баз данных в вычислительных системах. М., 1980.

Москович В. А. Информационные языки. М., 1971.

Солтон Дж. Динамические библиотечно-информационные системы. М., 1979.

Сэлтон Г. Автоматическая обработка, хранение и поиск информации. М., 1973.

Тиари Т., Фрей Дж. Проектирование структур баз данных. М., 1980.

Черный А. И. Введение в теорию информационного поиска. М., 1975.

АВТОМАТИЧЕСКИЙ (МАШИННЫЙ) ПЕРЕВОД

Автоматические системы, в которых воспроизводится тот или иной вид речевого поведения, относятся к классу кибернетических (функциональных) моделей.

С достаточной долей обобщения можно выделить три основных типа речевого поведения человека, для адекватного воспроизведения которых должны быть сконструированы, по крайней мере, три типа кибернетических моделей:

1) речевое поведение людей в процессе коммуникации, понимаемой здесь в широком смысле, — это не только непосредственное, но и опосредованное общение людей, например, при помощи печати, телевидения и т. п. Основной формой речевого поведения этого типа является диалог, моделируемый в различных экспертных системах;

2) речевое поведение в условиях обучения какой-либо отрасли знаний, в том числе и языку: этот тип поведения моделируется обучающими системами;

3) речевое поведение человека, переводящего тексты с неродного языка на родной и наоборот.

Очевидно, что системы машинного перевода (МП) являются кибернетическими моделями, воспроизводящими на ЭВМ именно третий тип речевого поведения.

Особенности реализации модели речевого поведения (речемыслительной деятельности) переводчика в виде системы МП определяются двумя классификационными параметрами:

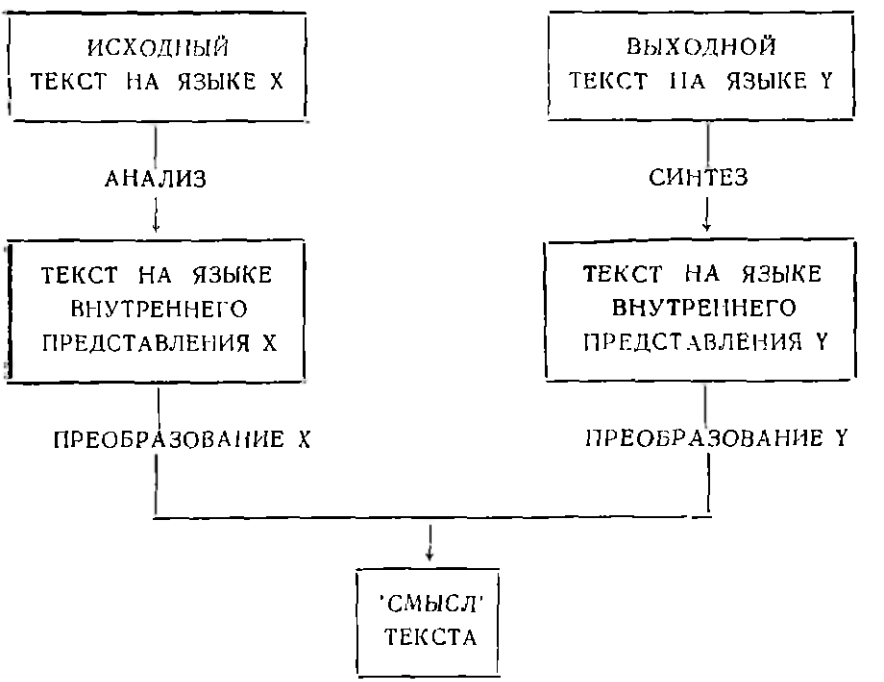
— степенью автоматизации (т. е. действительным объемом работы, который передается компьютеру). Здесь можно выделить ситуации (Кулагина, 1979, с. 3): 1) когда процесс перевода (почти) полностью реализуется ЭВМ; 2) когда переводит человек, а машина выполняет вспомогательную работу типа поиска в словаре;

— стратегией вовлечения человека — когда большая часть

работы выполняется компьютером, а человеком осуществляется пред-, интер- или постредактирование (Machine Translation, 1987, p. 12—13).

Специфика речевого поведения переводчика заключается в том, что исходным объектом является текст на одном языке, а результатом речевой деятельности — текст на другом языке, адекватный первому тексту по смыслу. Иными словами, этот результат должен быть обязательно оформлен в виде текста, в то время как при других типах речевого поведения такое оформление не носит обязательного характера — результатом речевой деятельности может быть, например, выполнение физических или умственных действий (в ответ на «команду» автора текста), накопление знаний для будущей деятельности и т. д.

Указанная специфика обусловливает определенную принципиальную структуру МП, содержащую обязательные компоненты независимо от того, какая конкретная схема системы АП объявлена ее создателями (см. рисунок).



Идеальная структура процесса АП.

Задачей анализа и преобразования X является «извлечение» смысла из текста на языке X на основе сведений из словаря, грамматики и правил употребления языка X, а задачей синтеза — конструирование или, как обычно говорят, синтези-

рование выходного текста на языке Y на основе смысла, преобразования Y, словаря, грамматики и правил употребления языка Y.

Обязательность компонентов анализа и синтеза, не столь явная при «человеческом» (ручном) переводе, очевидна при его автоматизации.

Необходимость преобразований можно установить даже в простейшем случае перевода, когда словарная единица одного языка сопоставима по значению со словарной единицей другого, а типы синтаксической структуры этих языков соответствуют друг другу вплоть до порядка слов. Очевидно, что в этом случае преобразование идет лишь по линии лексем, их грамматических форм и подстановки переводного эквивалента.

Однако такая идеальная ситуация, как известно, характерна лишь для ограниченной части словаря и синтаксиса языка и к тому же эффективно может быть определена лишь для каждой конкретной пары языков. Для абсолютного большинства лексических единиц (ЛЕ) и синтаксических конструкций двух языков такого однозначного соответствия установить не удается.

Одной из причин этого, как известно, является несовпадение объема значения многозначных слов в разных языках (ср. русское слово *стол,* которое соответствует английскому table лишь в своем значении предмета мебели).

Крайним случаем лексического несоответствия является так называемая безэквивалентная лексика какого-либо языка, как правило, описывающая предметы быта и культуры, не свойственные быту и культуре народа — носителя переводящего языка. Калькирование или описание лексического элемента в этом случае требует осмысления — а стало быть, и преобразования (к смыслу).

Как известно, зависимость словарных несоответствий разных языков от разного членения мира мышлением разных народов, отраженным в языковых фактах, устанавливает теория лингвистической относительности Сэпира — Уорфа (ср. одно понятие «снег» и одно слово для его обозначения в европейских языках, и множество понятий снега, которому соответствует множество наименований в языках некоторых северных народов).

Несоответствие между языками создается также из-за «тирании» узуса, употребления. В качестве примера можно указать на возможность образования будущего времени от английских глаголов persuade, win: I shall persuade, I shall win, и невозможность их «прямого», т. е. также при помощи форм будущего времени, перевода на русский ввиду отсутствия таковых у русских соответствующих глаголов: *убедить, победить,* — отсутствия, ни в коей мере не мотивированного семантикой данных глаголов (ср. неупотребительность, вызванную семантическими ограничениями: *колоситься, но *я колошусь).*

В качестве примера несоответствия структур можно привести рамочную конструкцию немецкого языка, образуемую разрывом компонентов сложного глагольного сказуемого и их особой позицией, не свойственной русскому и другим европейским языкам. Ср. Heute abends werde ich das Buch, das ich gestern gekauft habe, lesen beginnen и *Сегодня вечером я начну читать книгу, которую купил вчера.*

Если определить задачу перевода, в том числе автоматического перевода, с точки зрения показанных выше примеров несоответствия между языками, то можно сказать, что задачей перевода является приведение в соответствие по смыслу текстов двух языков, различные элементы которых (словарь, структура, порядок слов и т. п.) почти не соответствуют друг другу в аспекте способов и средств выражения содержания.

Представляется при этом, что приведение осуществляется при помощи указанных на схеме общих принципиальных, содержательных компонентов, которые должны присутствовать в любой системе МП независимо от объявленного устройства этой системы. Сама возможность приведения в соответствие и принципиальная переводимость с одного языка на другой связана с общими для всех языков свойствами: линейностью и знаковостью языка (Ф. де Соссюр), а в основе свойства знаковости (план содержания) лежит общность знаков мыслительной деятельности всех людей и общность физических и социальных явлений окружающего нас мира.

Как принято считать, причинами возникновения МП явился бурно растущий со второй половины XX в. поток информации на разных языках из разных стран и континентов, необходимость ее усвоения для научно-технического прогресса, недостаточность квалифицированных (в основном в конкретных областях знаний) переводчиков, высокая стоимость их подготовки и работы. Изложенное далее не следует рассматривать как обзор систем МП, даже краткий. Это, скорее, изложение некоторого представления о развитии МП.

Указанные социальные причины и вызвали необходимость поиска рациональных путей решения задачи перевода. Кроме этих, так сказать, «заставляющих» причин должны были иметь место и «разрешающие» условия. К первому из таких условий можно отнести особенности речевой деятельности переводчика на этапе чисто формального анализа структуры иноязычного текста (Панов, 1958), когда переводчик (очевидно, что речь идет о переводчике, выучившем иностранный язык, а не о билингве — человеке, владеющем им с детства как вторым родным языком), к примеру, имеет дело с текстом, специальная лексика которого ему неизвестна, известна лишь грамматика и функционально-грамматические элементы лексики типа союзов, предлогов, частиц, местоимений и др. Возможность формального анализа можно продемонстрировать на примере

знаменитой фразы Л. В. Щербы с придуманными лексемами и русской грамматикой «Глокая куздра штеко будланула бокра и курдячит бокренка». Читатель на основе этой фразы может сделать достаточное количество гипотез о ее синтаксисе и даже семантике (ср. *бокр — бокренок* и др.).

Именно формальные особенности указанного типа: от формы иноязычного текста к смыслу — явились причиной того, что объектом моделирования прежде всего стала деятельность переводчика, а не процессы понимания текста, написанного на родном языке пользователя, которые стали предметом моделирования гораздо позже, и именно на основании теоретического и практического опыта, накопленного разработчиками систем МП. Возможность формализации связана, таким образом, с необходимым «отчуждением», объективацией фактов языка при переводе иноязычных текстов, позволяющим лингвистам заменить и формально описать те из них, которые поддаются такому описанию без особого анализа семантики.

Официально признанным началом работ по автоматическому переводу считается 1949 г., и это связано со вторым, важнейшим, «разрешающим» условием развития автоматического перевода — а именно, созданием электронно-вычислительной техники. Однако сама идея МП при помощи механизмов возникла гораздо раньше. Существуют сведения (заметка в газете «Yaba maa» от 24 февраля 1924 г.) об изобретении А. Кахером пишущей машины-переводчика (Переводная машина в Таллине, 1962). В 1933 г. в Советском Союзе был выдан патент П. П. Смирнову-Троянскому на переводную машину с одного языка на другие (Переводная машина П. П. Троянского, 1959).

Лингвистический принцип перевода заключался в сопоставлении эквивалентных по смыслу ЛЕ двух и более языков. Автоматизм этого сопоставления обеспечивался механическим способом. Отсутствие электронно-вычислительной техники, а также замкнутость системы патентного дела в нашей стране в то время явились причиной того, что об этой работе в мировой науке стало известно лишь в 50-х годах XX в., т. е. в период создания систем МП с помощью ЭВМ. Впрочем, та же участь постигла запатентованное в 1933 г. изобретение машины для перевода: «Механический мозг» французского инженера Георгия Арцруни (Corbe, 1960).

Как уже говорилось выше, решающим фактором возникновения современных систем МП явилось создание в 40-х годах XX столетия электронно-вычислительной техники, способной выполнять разнообразные операции с большой скоростью. Первые системы МП возникли в США. Исторически сложилось так, что в самом начале общий лингвистический подход к построению систем МП был воспринят из теории и практики криптографии.

Во время второй мировой войны в разных странах, в особенности в Англии и США, были достигнуты большие успехи не только в «ручной», но и в машинной дешифровке сообщений — так, например, была создана дешифровальная машина «Ультра», позволившая расшифровывать приказы и сообщения гитлеровского командования. Дешифровка закодированных сообщений с лингвистической точки зрения представляет собой процесс перевода текста, записанного на неизвестном языке, с неизвестной лексикой и по неустановленным грамматическим правилам. Такой перевод является процедурой перехода от одной кодовой системы к другой.

При переводе с естественного языка задача, на взгляд специалистов-дешифровщиков, была проще: заранее известны особенности языков, их лексика, морфология и синтаксис, описанные для большинства современных языков в огромном количестве словарей и грамматик. Вычислительная техника давала возможность расположить в своей памяти и словари, и грамматики.

В 1946—1947 гг. идея МП при помощи электронно-вычислительной цифровой машины обсуждается как проблема в Принстонском научно-исследовательском институте (У. Уивер, А. Л. Бут и др.). Первое концентрированное выражение эта идея нашла в меморандуме американского специалиста по дешифровке Уоррена Уивера, разославшего 15 июля 1949 г. свой меморандум двумстам специалистам в области лингвистики, вычислительной техники и дешифровки.

У. Уивер связывал принципиальную возможность МП с наличием сходных черт между совершенно различными языками, а также с наличием у каждого языка формальных характеристик, которые можно учитывать до включения в процесс перевода. Вообще, по мнению У. Уивера, книга, написанная по-китайски, — не что иное, как книга, написанная по-английски и закодированная китайским шифром (Машинный перевод, 1957, с. 33—45).

От меморандума Уивера и идет отсчет времени развития МП как отдельного научного направления. Прошедшие сорок пять лет его развития позволяют подвести некоторые итоги и проанализировать пройденный путь. Весь этот путь можно разделить на три основных периода (существует и другая периодизация, см., напр.: Hutchins, 1986; Кулагина, 1979; Марчук, 1983, с. 91; Апресян и др., 1989, с. 4—5). Хронология этой периодизации весьма условна. Попытки установить хронологические рамки периодизации малоубедительны.

Представляется, что на фоне условного хронологического деления (например, такого: 1-й период: 1949 — середина 60-х годов; 2-й период: середина 60-х — конец 70-х; 3-й период: 80-е годы нашего века) принципиально существенным основанием деления должны служить научно-технические результаты. По-

этому часто говорят о системах МП первого, второго и третьего (современного) поколения, имея в виду развитость лингвистического, математического и машинного обеспечения, а также практические результаты. Не имеет смысла также производить мелкую хронологическую периодизацию, как это сделал М. Хатчинс, ввиду условности даже крупных хронологических рамок.

Первый период, провозглашенный меморандумом Уивера, начался в ноябре 1949 г. созданием первого центра в Вашингтонском университете (система русско-английского МП), затем в 1950 г. под руководством А. Д. Бута был создан центр французско-английского перевода. В Советском Союзе первый центр перевода появился в 1955 г. в Москве в Математическом институте имени В. А. Стеклова АН СССР под руководством О. С. Кулагиной.

В этот период были получены первые обнадеживающие результаты и проведены первые машинные эксперименты. Так, в 1954 г. в Джорджтаунском университете в США был продемонстрирован МП с русского языка на английский с помощью системы GAT, разработанной под руководством проф. Леона Е. Достерта. Это был первый опыт перевода с одного языка на другой, сделанный при помощи ЭВМ (IBM-701), известный в науке под именем Джорджтаунского эксперимента (Машинный перевод, 1957, с. 171). В этом эксперименте была опробована процедура «прямого» МП (к этому типу относятся системы, спроектированные для одной языковой пары), словарь системы включал в себя всего 250 слов, а грамматика содержала 6 синтаксических правил (Machine Translation, 1987, p. 22—23). В СССР первая экспериментальная система МП с французского языка на русский была создана в 1955—1956 гг. (авторы О. С. Кулагина, И. А. Мельчук) (Кулагина, Мельчук, 1956). В этой системе словарь основ имел объем в 1236 слов, был введен словарь оборотов в 250 единиц. Единицей перевода являлось предложение. Алгоритм морфологического анализа на основе работы с таблицами окончаний приписывал словам переводимой фразы необходимую грамматическую информацию. «Продвинутость» этой системы заключалась также в наличии блока снятия омонимии. В блоке синтеза русской фразы предусматривалось указание на правильный порядок слов. В то же время была спроектирована первая система англо-русского перевода (рук. Т. Н. Молошная) (Молошная, 1960).

Бурный рост нового направления отразился в большом количестве представительных конференций, симпозиумов, публикаций, а также в возникновении огромного количества центров, групп и лабораторий, занимающихся проблемой МП. К началу 60-х годов зафиксировано следующее их распределение по количеству и странам: ФРГ — 22, США — 21,

СССР — 16, КНР и Япония — по 6, Франция — 5, Чехословакия — 4, Великобритания и Венгрия — по 3, ГДР и Италия — по 2. Возникли также группы в Бельгии, Болгарии, Израиле, Канаде, Люксембурге, Мексике, Польше, Румынии, Финляндии, Югославии (Автоматический перевод, 1967, с. 509).

В целом следует отметить, что начало работ по МП характеризовалось общим подъемом — лингвисты открыли новое направление исследований, программисты поставили сложные и интересные задачи. В этот период казалось, что вся информация о языках и о процессе перевода есть, ее нужно только формализовать и цель будет достигнута — будет создана универсальная система МП. Однако первые и последующие опыты и выявленные в них трудности перевода показали, что не все так просто.

В то же время эти работы стимулировали исследования не только в области МП, но и в области теоретической и структурной лингвистики.

Первый период развития автоматического перевода в целом был необычайно плодотворным. В этот период разрабатывалось и наметилось большинство фундаментальных проблем: построения автоматических словарей, создания языка-посредника (ЯП), формализации грамматики языков, разрешения омонимии слова и синтаксических конструкций, обработки фразеологизмов и многие другие.

В это же время были разработаны первые варианты алгоритмов синтаксического анализа, остающиеся базовыми и сегодня: анализ на основе учета синтаксических зависимостей и непосредственно составляющих, предсказуемостный анализ, многовариантный синтаксический анализ, анализ методом «фулькрумов», конфигурационный анализ и др.

Анализ в терминологии моделей зависимостей и моделей непосредственно составляющих предполагает установление структуры связей целого предложения. Для снятия многозначности такой структуры использовались методы различных фильтров (см., например, работы лаборатории математической лингвистики ЛГУ, рук. С. Я. Фитиалов, Г. С. Цейтин; Гренобльской группы GETA — Centre d'études pour la traduction automatique, рук. В. Вокуа). В качестве «сильного» фильтра Д. Е. Хейсом (The RAND Corporation Automatic Language Data Processing Group. Santa Monica (Calif.)) было предложено понятие «силы» связи (value). Предпочтительным оказывался тот вариант анализа, «сила» связей которого была выше (Автоматический перевод, 1971, с. 41).

Предсказуемостный анализ (Национальное бюро стандартов США, И. Родес; Вычислительная лаборатория Гарвардского университета (Масс.), рук. А. Эттингер) основывается на особой структуре словарных статей, где каждому слову приписывалась информация обо всех его возможных свя-

зях. Процедура анализа основывается на просмотре текста слева направо, при этом для каждого очередного слова проверяется, является ли оно по тем или иным параметрам предсказанным предыдущим. Если нет, то поиск «хозяина» осуществляется левее. В группе И. Родес предусматривался одновариантный анализ — первое же слово слева от анализируемого, соответствующее ему по предсказаниям, считалось «хозяином» и, следовательно, любому предложению ставилась в соответствие только одна синтаксическая структура.

Многовариантный анализ, дающий возможность получить несколько вариантов синтаксической структуры предложения (С. Куно, А. Эттингер, Вычислительная лаборатория Гарвардского университета), был шагом вперед. Это связано с тем, что многовариантность основана на естественной множественности синтаксического потенциала слова предсказывать синтаксически связанные с ним слова. Первая грамматика многовариантного анализа содержала 2100 правил и 82 синтаксических предсказания. Диаграмма структуры предложения строилась по модели непосредственно составляющих, при этом перебирались все возможные варианты, большое число которых сокращалось при помощи различных фильтров. В алгоритмах предсказуемостного анализа многовариатность—не столько следствие «двусмысленности», объективно присущей синтаксическому строю фразы, сколько следствие избранного формализма перебора вариантов без достаточной разработанности формальных и лингвистических ограничений. В этом заключается слабость этих систем, хотя сама идея синтаксических предсказаний плодотворна и лежит в основе некоторых современных систем автоматического анализа (Цейтин и др., 1982), проектируемых на основе грамматики предсказаний (Откупщикова, 1982).

Следует отметить, что это противопоставление одновариантного и многовариантного анализа, выкристаллизовавшееся еще в первый период МП, остается необычайно актуальным. При одновариантном анализе почти «наугад» выбранная структура предложения может оказаться неверной, а возможности вернуться к началу анализа такие системы не имеют. При многовариантном анализе нам грозит кризис размерности: количество построенных структур может быть существенно бо́льшим, чем возможности выбора правильного варианта.

В то же время очевидно, что хороший переводчик (или переводящая система), понимая весь текст, в который входит данная фраза, «выдает» один вариант ее перевода и притом правильный — многозначность фраз «разрешается» в связном тексте. Однако при этом широко используются семантические сведения, сведения об описываемой текстом ситуации действительности (так называемые нелингвистические знания) и, наконец, сведения о синтаксических и семантических закономер-

ностях целого текста. В системах и первого, и второго периода это не использовалось совсем или использовалось фрагментарно. Основной причиной такого положения в то время можно считать отсутствие разработанного в указанных направлениях лингвистического обеспечения.

Конфигурационный анализ (кафедра математической лингвистики ЛГУ (Белопольская, Ковригин, 1963; Воронин, 1961), кафедра структурной и прикладной лингвистики МГУ (см., напр., Дрейзин, Рашитов, 1961), группа автоматического перевода ИЯ АН СССР (Молошная, 1960) и др.) основывается на анализе входного текста путем сравнения с набором предварительно выделенных синтаксических конструкций (конфигураций, формул) переводимого языка. При совпадении установленной конфигурации входного текста с одной из конфигураций набора текстовая конфигурация считается распознанной, «свертывается» и в дальнейшем анализе фигурирует в свернутом виде. При синтезе полученная из анализа конфигурация сравнивается с соответствующей ей конфигурацией переводящего языка. Соответствие конфигураций установлено в системе предварительно.

Анализ методом «фулькрумов» (группа автоматического перевода фирмы Bunker Ramo-Woldridge Inc, Canoga Park, Calif., рук. П. Гарвин) основывается на том, что предложение разбивается на компоненты низшего порядка (блоки), которые объединяются в более крупные фрагменты при помощи серии последовательных просмотров (циклических прогонов) фразы. При этом на этапе каждого очередного прогона устанавливаются опорные точки анализа (fulcra), которые являются главными, «управляющими» словами синтаксических конструкций (блоков), установление структуры и границ которых и является целью циклического прогона.

Цепочечный анализ (З. Харрис, Н. Сейгер) предполагает распознавание синтаксической структуры фразы путем разбиения цепочки синтаксических категорий, при помощи которой представлено исходное предложение, на элементарные цепочки, напр.: NtV=power corrupts 'Власть развращает' — это элементарная цепочка.

Корреляционный анализ (система «Мульмистор», С. Чекатто, Э. Ф. Главерофельн) заключается в том, что слову приписывается набор индексов, каждый из которых отражает способность данного слова коррелировать с другим словом, имеющим идентичный индекс. Высокое число корреляций при анализе снижается при помощи лингвистических ограничений, например, на порядок слов, семантику и т. п.

В то же время исследователи начинают разработку проблем анализа целого текста для задачи МП, в частности, составляются правила соотнесения их с антецедентами-сущест-

вительными (Шумилина, 1961; см. также: Автоматический перевод, 1967, с. 235 и сл.).

Системы МП, проектируемые в это время, делятся на два класса: бинарные, осуществляющие перевод с одного языка на другой, и множественные (с ряда языков на один или несколько).

Превалирование систем первого типа в то время и в последующие годы обусловлено естественностью ситуации, моделирующей человеческий перевод, а также большей теоретической и практической возможностью установления бинарных соответствий фактов каждой пары языков. Недостатком бинарных систем МП являлось то, что они содержали в себе жесткую зависимость анализа от свойств входного, переводимого языка, которая усугублялась жесткостью самих алгоритмов перевода: факты языка в них не были отделены от детерминированности хода анализа.

Множественный перевод позволял избежать таких жестких зависимостей и давал возможность создавать независимые алгоритмы анализа и синтеза. При множественном переводе, понимаемом как сумма бинарных переводов, количество их вычисляется по формуле $n(n-1)$, где n — множество языков, участвующих в процессе перевода.

Первые разработчики множественных систем предложили использовать в качестве инварианта для анализа и синтеза язык-посредник (интерлингву) (Машинный перевод, 1957).

Первая значимая разработка идеи языка-посредника (ЯП) при множественном переводе принадлежит Н. Д. Андрееву (Андреев, 1957). В дальнейшем появился ряд работ сотрудников ЭЛМП (экспериментальной лаборатории машинного перевода — Ленинград, ЛГУ, рук. Н. Д. Андреев), на базе которой затем была организована лаборатория математической лингвистики (рук. С. Я. Фитиалов, затем Г. С. Цейтин), а также сотрудников других групп и центров в Советском Союзе и за рубежом, в которых обсуждалась идея ЯП.

Согласно концепции Н. Д. Андреева в основе понятия ЯП лежит представление об отсутствии конгруэнтности естественных языков, а стало быть, о закономерной искусственности ЯП, из которого должны быть удалены неконгруэнтные элементы (типа артиклей, арабских глагольных видов и т. п.) и сохранены конгруэнтные для большинства языков элементы типа трех глагольных времен, множественного числа и т. п. Таким образом, в первых работах, посвященных разработкам такого рода, ЯП по своей структуре был близок естественному языку и описывался в его терминах.

Недостаточная разработанность семантической теории для прикладных задач не давала лингвистам в то время возможности отойти от обычных лингвистических представлений, хотя всегда имелось в виду обобщение свойств естественных язы-

ков, выявляемое их сопоставлением, а само сопоставление строилось на основе учета грамматических значений, грамматических категорий и общих семантических значений у ЛЕ разных языков (Мельчук, 1959; Автоматический перевод, 1967). В ЭЛМП был проведен микроэксперимент перевода с помощью ЭВМ с 18 языков на русский, целью которого было продемонстрировать идею ЯП.

Идея множественного перевода через ЯП, однако, не получила должного развития в дальнейшей истории МП, несмотря на неоднократные последующие попытки ее воплощения. В теоретическом плане это связано с тем, что ЯП описывался в терминах грамматики и лексики, хотя и достаточно обобщенных, в то время как он, по идее, должен быть сугубо семантическим, должен быть описан при помощи адекватного семантического представления.

Отмеченная критиками низкая адекватность перевода через ЯП оказалась связанной с потерей ряда существенных характеристик переводимых текстов: коммуникативных и прагматических установок автора текста, актуального членения и т. п., — неизбежной при предлагаемых моделях ЯП.

С точки зрения практики лингвистических разработок систем МП нереализуемость ЯП была обусловлена необычайной усложненностью этапов преобразования, а в техническом отношении — низкими техническими показателями объемов памяти и быстродействия у ЭВМ первого поколения, что затруднило построение впечатляющих экспериментальных систем даже для бинарного перевода.

В настоящее время ряд указанных преград для реализации ЯП преодолен. Новый виток интереса к ЯП связан с созданием систем искусственного интеллекта, ведь именно от этой идеи можно протянуть непосредственную связь к интеллектуальным системам понимания текста, разрабатываемым современной кибернетикой. Однако и сегодня МП с использованием ЯП является тем идеалом, стремление к которому разделяют создатели любых систем. Следует отметить в то же время, что разработка ЯП имела большое значение для развития семантической теории, семантического представления фраз и текстов естественного языка.

Несомненный вклад концепции ЯП в развитие МП заключается также в признании независимости анализа и синтеза, до определенной степени опирающихся лишь на свойства переводимого и переводящего языков; признание необходимости разработки унифицированного и универсального представления смысла переводимого предложения, что оказывается релевантным и для высококачественного бинарного перевода.

Несмотря на активную работу в области создания систем МП в первый период разработчики не пришли к желаемым практическим результатам — созданию систем промышленного

перевода, удовлетворяющего запросы пользователя. Это имело большие практические последствия. Часть исследователей отошла от решения проблем МП. Определенную роль в свертывании работ ряда групп сыграла экстраполяция выводов отчета специальной комиссии Национальной Академии наук США (ALPAC report), сделанного на основе анализа системы МП ВВС США, на все разработки систем МП, даже более продвинутые. В результате (хотя системы и ВВС США, и Джорджтаунская продолжали действовать) у многих коллективов, к тому времени начавших создавать системы более продвинутого типа, возникли организационные и финансовые трудности (Кулагина, 1979, с. 14).

Не последнюю роль сыграл в этом плане также скепсис некоторых ученых, сомнения в возможности воплощения идеи МП. Этот «новый» скепсис относился к возможности МП научно-технических текстов. Что же касается текстов художественной литературы, особенно поэзии, то большинство создателей практических систем МП с самого начала намеренно исключили их из переводимого материала ввиду их образности, использовании нарушений языковых закономерностей в качестве художественных приемов (ср. нарушение порядка слов, проективности и т. п.).

Особенно известны в этом отношении взгляды И. Бар-Хиллела, который считал, что высококачественный полностью автоматизированный перевод невозможен, потому что существует многозначность, которую нельзя снять простым обращением к энциклопедическим знаниям о внеязыковом мире (ср. the box was in the pen — *ящик был в (детском) манеже,* но **ящик был в пере*). По мнению И. Бар-Хиллела, невозможно также точно перевести словосочетания типа slow neutrons and protons по той причине, что в самом языке нет никаких формальных показателей разрешения этой двусмысленности — нужны знания о мире. Поэтому полная автоматизация перевода, по мнению И. Бар-Хиллела, является утопией (Бар-Хиллел, 1962).

Все вышесказанное и повлекло за собой действия организаторов науки во всем мире по сокращению числа групп, занимающихся данной проблемой, и к концу второго периода из 100 групп осталось 38 (Кулагина, 1979, с. 16—18).

Однако, несмотря на то, что к этому времени у многих ученых накопилось достаточно скепсиса и разочарования, наиболее последовательные и упорные не сложили оружия. Начался второй этап МП — углубленные теоретические исследования, создание формализованных лингвистических моделей, системных и функциональных грамматик. Особое внимание уделялось в это время моделированию семантики.

Этот период МП без машин, с одной стороны, определил необходимость специальных исследований языкового материала для создания частных моделей перевода, с другой — позволил

подвести итоги первого периода и определить возможность получения результатов при переходе от глобальных задач к созданию прагматически ориентированных систем.

Начатые в первый период разработки моделей формальных грамматик: теоретико-множественной модели, моделей грамматик зависимостей, непосредственно составляющих, моделей порождающей грамматики и семантики и др. — во второй период МП продолжают разрабатываться, интерпретируются, применяются к более или менее полным описаниям грамматических структур отдельных языков, а также апробируются в более фундаментальных алгоритмах автоматического анализа.

Ко второму периоду относится также более интенсивная разработка принципов семантического анализа и осознание необходимости привлекать семантику в системах анализа для получения высококачественного перевода — в одних системах путем введения особого семантического уровня (уровней) анализа (например, ФРАП), в других — при помощи заданного широкого списка типов синтаксических отношений (например, в модели «смысл — текст»), носящих семантический характер, в третьих — заданием смыслового инварианта типа ЯП и семантических правил приведения к нему (ср. этапы интерпретации в системе ЛГУ) и т. п. В то же время второй период характеризуется продолжением работ по созданию, а также экспериментальному и промышленному опробованию систем МП.

Путем увеличения словаря и базовых алгоритмов развивалась Джорджтаунская система (GAT) и ее дериваты, в частности система SYSTRAN. В 1976 г. англо-французская версия этой системы была передана Европейскому Экономическому Сообществу. В этой системе предусмотрен синтаксический анализ, устанавливающий основные синтаксические связи, обрабатывающий однородные конструкции, неоднозначности и т. п. (Марчук, 1985), а также процедура трансфера (преобразования). *Анализ и синтез достаточно независимы и могут быть аппликированы к системам с другими языками-партнерами.*

Гренобльская система русско-французского перевода (GETA, рук. Б. Вокуа) в усовершенствованном варианте, разрабатываемом с 1971 г., предусматривала получение (в результате анализа) глубинных синтаксических структур в логической форме (при помощи аргументов и предикатов), из которых в результате ряда трансформаций при помощи семантических, синтаксических и морфологических данных (и данных словаря) получается поверхностная форма выходного языка. Анализ многоуровневый, организован в виде последовательности 11 блоков обработки — например, приведения аналитических форм к синтетическим, сегментирования фразы, анализа именных групп и т. д. вплоть до поиска антецедентов и определения синтаксической и логической функций восстановленных узлов. Эта система имеет развитое программное обеспечение ARIANE, в частности, она

предусматривает взаимодействие системы и пользователя с помощью базы данных, а также использование развитых средств управления процессом, обработки текста (применение параллельной обработки, эвристик и т. п.).

В системе TAUM (Монреальский университет, Канада, рук. Р. Китредж) осуществляется англо-французский перевод на уровне поверхностного синтеза (Синтаксический компонент . . ., 1981). Система содержит блок трансфера, задачами которого являются лексическое и структурное преобразование: перевод лексем, преобразование некоторых частей дерева составляющих для облегчения синтеза французской фразы, перевод глагольных времен и модели управления. Языком программирования является Q (Quebec) — язык, построенный А. Кольмером (Синтаксический компонент. . ., 1981, с. 92). Комбинаторность и гибкость анализатора в этой системе (хранение первоначальных гипотез анализа до его конца, установление порядка применения правил несколькими способами и т. п.) создают возможность ориентации перевода на поверхностный синтаксис.

Система TAUM активно развивалась с 1968 по 1980 г., с 1977 г. она функционирует как промышленная система, переводя в год тексты объемом 5 млн. слов (Machine Translation, 1987, p. 12).

В группе алгебраической и вычислительной лингвистики Карлова университета (Чехословакия, рук. П. Сгалл) (Синтаксический компонент . . ., 1981, с. 84) разработана система англо-чешского МП простых предложений (на ограниченном словаре). Процесс анализа разделен на четыре основных этапа: доморфологический анализ, морфологический анализ, синтаксический анализ, перевод. Синтаксический анализ построен по типу расширения области действия синтаксических правил: вначале анализируются единицы узкого контекста и т. д. вплоть до контеста, выходящего за рамки предложения. Таким образом, это типичный случай анализа «снизу вверх». Синтаксический анализ в связи с этим делится на несколько этапов. В систему введены «аварийная» грамматика для анализа нерегулярностей синтаксиса, правила восстановления элементов, отсутствующих в предложении, но необходимых для установления его смысла, широко используется семантическая интерпретация синтаксических отношений.

Система SUSY (Саарбрюкенн, Саарский университет, рук. Г. Элере, Х. Д. Маас) разрабатывается примерно с 1970 г. (Синтаксический компонент . . ., 1981, с. 90). Процесс перевода состоит из предварительной подготовки данных, анализа, преобразования (собственно перевода) и синтеза. Алгоритмы отделены от грамматик конкретного языка, поэтому подключение нового языка к системе не требует создания большого числа новых программ. Разрабатываются немецкий, русский, английский и французский анализы и немецкий синтез. Результатом анали-

за являются «базисные структуры», имеющие вид деревьев порождения. Система словарей (словари анализа, семантические словари и словари перевода) позволяет использовать любую информацию в процессе анализа, что особенно важно для семантической информации. Принцип анализа — комбинированный: первая часть анализа начинает обработку «сверху вниз», вторая — «снизу вверх», третья «склеивает» те и другие результаты. Для отбраковки неправильных синтаксических структур в конце анализа предусмотрено подключение процедур семантического анализа.

Несмотря на сокращение групп, занимавшихся МП в нашей стране, рассматриваемый нами второй период характеризуется достаточной плодотворностью разработок, выражающейся:

а) *в создании улучшенных версий ранее разрабатывавшихся* систем в аспекте как лингвистического, так и математического обеспечения;

б) в разработке новых систем (Всесоюзный центр переводов, группа «Статистики речи» в ЛГПИ им. А. И. Герцена);

в) в интенсивных разработках систем автоматического анализа (а не полных систем перевода);

г) в доведении систем МП до экспериментальной проверки разной значимости.

Последний пункт совмещает в себе достоинства и недостатки разрабатывавшихся в это время систем: имея экспериментальную апробацию, ни одна из них практически не была доведена в этот период до промышленной эксплуатации. В большой степени это объясняется недостаточным техническим и организационным оснащением коллективов, занимавшихся разработкой систем МП.

В Институте прикладной математики АН СССР в этот период была создана вторая, усовершенствованная версия системы французско-русского перевода — ФР-П (рук. О. С. Кулагина) (Кулагина, 1979), предназначенная для перевода французских математических текстов. Система разрабатывалась с 1967 по 1976 г. Этап преобразования (собственно перевода, *трансфера) на входе имеет поверхностную синтаксическую* структуру французских фраз (перевод пофразный), а на выходе — поверхностную синтаксическую структуру соответствующих по смыслу русских фраз. В основе этого подхода лежит гипотеза о сходстве слов двух языков по значению и фраз по значению и строению. Целью анализа является построение множества синтаксических структур фразы в виде расположенных деревьев зависимостей. Выделено 52 типа синтаксических связей.

Анализ многовариантный, с набором правил-фильтров, которые учитывают силу и дальность синтаксической связи, совместность/несовместность связей, проективность/непроективность структур. Введена иерархия фильтров, соответствующая по-

рядку их применения (вначале применяются простые, общие фильтры, затем сложные, частные). Фильтры позволяют из набора гипотез составить допустимые, которые лежат в основе создания одного из вариантов многовариантного анализа (Кулагина, 1979; Синтаксический компонент . . ., 1981).

Если основной принцип лингвистического подхода, используемого в системе ФР-П, можно считать исключающим идею ЯП, то модель синтаксического анализа, известная под названием модель «смысл — текст» (рук. И. А. Мельчук), напротив, представляет собой попытку последовательного развития идеи представления смысла переводимой фразы, которое является целью анализа и исходным материалом для синтеза. При этом одно и то же представление смысла соответствует множеству «поверхностных» языковых выражений. Например, смысл 'мальчик' 'читает' 'книгу' соответствует набору языковых выражений: *Мальчик читает книгу, Книга читается мальчиком, Мальчик, читающий книгу, Книга, читаемая мальчиком, Чтение книги мальчиком, Чтение книги, которое осуществляется мальчиком* и т. д.

При разработке русского синтаксического анализа используются идеи конфигурационного анализа, предложенные в первый период развития МП. Выделено 31 отношение непосредственной доминации (ОНД), которые содержательно характеризуют многообразие типов синтаксических отношений слов в предложении. Целью синтаксического анализа является установление синтаксических связей между словами переводимого предложения, приписывание этим связям необходимых ОНД и приведение в соответствие заданному тексту особой характеристики, представляющей смысл этого текста. Система автоматического анализа имела как прикладное значение (для автоматического перевода, реферирования, поиска информации), так и теоретическое, так как «алгоритм анализа является действующей и экспериментально контролируемой моделью „анализирующей части языка"» (Автоматический перевод, 1967, с. 226; Мельчук, 1974).

В лаборатории математической лингвистики НИИММ ЛГУ до стадии экспериментальной проверки была доведена система англо-русского перевода (Лейкина и др., 1966). В ней предусматривалась независимость анализа и синтеза, задачей преобразования и интерпретации являлось приведение к унифицированному представлению. Анализ по своему характеру — пофразный, множественный. Лингвистические сведения, сформулированные в рамках модели грамматики зависимостей, отделены от процедур (алгоритмов) обработки этих сведений. Эффективность работы алгоритма обеспечивалась выполнением условий на правильный синтаксический анализ: это условия проективности, древесности, обязательности связей, совместимости и т. п. Кроме того, в лаборатории несколько позже на-

чала создаваться и экспериментально опробоваться система русского синтаксического анализа с учетом предпочтения и порядка замыкания связей (Кремнева и др., 1976).

Под руководством Ю. Д. Апресяна («Информэлектро») была создана система французско-русского перевода (ИНФЭЛ) с мощным лингвистическим обеспечением, в котором учитывались эллипсисы и сильная непроективность (Синтаксический компонент..., 1981). Перевод осуществлялся на уровне поверхностных синтаксических структур. Принципы построения ИНФ-ЭЛ следующие:

1) согласованность описания грамматик переводимых языков на основе единой модели «смысл — текст»;

2) обратимость правил анализа и синтеза на основе достаточности информации в них;

3) отделенность грамматических и словарных описаний (кроме правил собственно перевода) от алгоритмов и тем самым от программ;

4) независимость описания переводимых языков;

5) универсальность в смысле ориентации на разные предметные области;

6) создание формального аппарата (модели и языка) для представления лингвистических знаний.

Первая очередь системы была реализована на ЭВМ «Ирис-50».

В МГПИЯ им. М. Тореза создавался многаспектный англо-русский словарь для системы МП АРМАС, статьи которого содержали разнообразные лингвистические характеристики словарной единицы: морфологические, синтаксические, семантические, сведения о лексических функциях, о модели управления, о переводных эквивалентах и т. д. (Шаляпина, 1974).

Этот словарь входит важнейшей частью в лингвистическое обеспечение системы англо-русского автоматического перевода — АРАП (рук. З. М. Шаляпина). Принципы, разработанные для этой системы, применены и в системе японско-русского автоматического перевода — ЯРАП (рук. З. М. Шаляпина).

Для этих систем характерно наличие одного синтаксического уровня анализа — семантико-синтаксического. Для записи семантической сочетаемости языковых единиц разработан специальный язык. Вся основная лингвистическая информация представлена в однотипных словарных статьях, в которых словарные единицы представлены в их грамматических, лексических, сочетаемостных и других признаках, необходимых и достаточных для пофразного перевода.

Во Всесоюзном центре переводов во второй период создавалось несколько систем перевода. Отметим две из них: ФРАП — система французско-русского автоматического перевода (рук. Н. Н. Леонтьева) (Леонтьева, 1987) и АМПАР —

система англо-русского перевода (рук. Ю. Н. Марчук) (Марчук, 1983).

Система ФРАП представляет собой интересную попытку моделирования действий переводчика, для которого переводимый язык не является родным, а переводимый текст содержит элементы неизвестного (на данном этапе перевода). Научно-исследовательской целью при этом является выяснение «силы» чисто грамматического подхода. В связи с этим процесс перевода разбит на большое количество блоков, каждый из которых соответствует определенному уровню анализа переводимой фразы, и лишь использовав все формально-грамматические средства, процедура перевода «предлагает» включение семантических уровней анализа, которых, так же как и грамматических, предусмотрено несколько. Уровни анализа называются представлениями. Концепция перевода в системе ФРАП доказывает «силу» синтаксического компонента системы. Особенностью системы является развитый семантический аппарат, обращение к которому предусмотрено после работы первого синтаксического блока. Экспериментальной проверке подвергались подсистемы разных уровней.

Система АМПАР также многоуровневая, ориентированная на грамматический анализ и развитый лексический анализ: отдельными блоками обрабатываются обороты, однозначные и многозначные слова. Преобразование поэтому касается лишь структуры предложения. Система имела практическое применение.

С 1970 г. начата разработка систем МП во Всесоюзной группе «Статистика речи» (рук. Р. Г. Пиотровский). В основу этих систем положен итеративно-стратификационный подход, остающийся актуальным и сегодня. При этом подходе задача получения непротиворечивого описания входного текста, во-первых, решается поэтапно, а во-вторых, необязательно должна решаться до конца. Создаваемая с 1975 г. в лаборатории инженерной лингвистики ЛГПИ им. А. И. Герцена система СИЛОД разрабатывалась как многоязычная система с мощным словарным и морфологическим обеспечением и грамматикой, использующей трансфер. В этот период были созданы и экспериментально опробованы версии англо-русского и французско-русского МП, заложены основы фреймового и тезаурусного анализа, позволяющего распознавать структуру и значение отдельных синтаксических конструкций (Пиотровский, 1979; Чижаковский, Беляева, 1983). В системе предусмотрено уровневое построение алгоритмов МП, являющееся, по сути, развитием метода «фулькрумов», предложенного на первом этапе МП.

Краткое рассмотрение систем второго поколения у нас в стране и за рубежом может представить их основные харак-

теристики. Отметим следующие общие особенности этих систем (Кулагина, 1979; Hutchins, 1986):

1) многоуровневость анализа и при этом хорошая представленность синтаксического уровня;

2) многовариантность результатов анализа, что, как уже отмечалось выше, являло собою прогресс, так как многовариантность в этом случае повышает вероятность появления среди множества вариантов одного правильного, в то время как жесткий одновариантный анализ не может всегда гарантировать правильность выбранного варианта;

3) переход от жесткого алгоритма, включающего в себя все лингвистические сведения, что при добавлении или изменении каких-либо лингвистических правил или признаков влекло за собой переделку алгоритма, к более гибкой системе, в которой все лингвистические сведения отделены от механизма их обработки, т. е. от алгоритма (Лейкина и др., 1986; Кулагина, 1979).

Это имело важные следствия, прежде всего, для организации работ, так как усилия лингвистов и математиков могли быть направлены на более интенсивную разработку специфических аспектов системы. Кроме того, это вызвало необходимость разработки и применения формальных моделей языка, которые позволили (в пределах заданных формализмов) унифицированным способом описывать грамматику языка. Именно ко второму периоду развития МП относится применение формальных моделей для описания разных языков, а также разработка новых моделей, имеющих широкие возможности для более адекватного описания языковых фактов.

Таким образом, второй период развития МП характеризовался как развитием практических систем первого поколения, накоплением опыта их эксплуатации и расширением словарной базы, так и созданием более оснащенных теоретических систем, хотя они имели в основном лишь экспериментальное опробование.

Полученные практические и теоретические результаты определили готовность перехода к третьему периоду в развитии МП. Однако этот переход был предопределен еще и рядом экстралингвистических фактов.

Во-первых, к концу 70-х годов необычайной интенсивности достиг поток информации. Так, например, американскими и канадскими корпорациями переводится до 2,5 млрд. слов в год. Если затраты в США на перевод в 1966 г. составляли 22 млн. долларов в год, то в 1981 г. они достигли уж 200 млн. долларов. В нашей стране только на оплату заказных переводов (не считая оплаты штатных переводчиков) тратятся огромные средства. Однако объем потоков информации (научно-технических текстов, патентов, документации и спецификаций) настолько велик, что силами переводчиков просто не может быть

обработан. Следовательно, интенсификация труда технических переводчиков на основе автоматизации процесса перевода стала в настоящее время насущной необходмостью.

Во-вторых, опыт создания и использования в издательской и лексикографической работе промышленных лингвистических процессоров типа спеллеров, систем обработки слов (word processors), мощных терминологических банков определил возможность получения практических результатов от применения «лингвистических» автоматизированных систем.

И наконец, в-третьих, особые предпосылки для перехода к новому этапу МП определило развитие самой вычислительной техники — появление дисплеев, персональных компьютеров, сканнеров, рост мощности персональных ЭВМ; все это изменило подход к теоретико-лингвистическим исследованиям и перенесло их в практическую плоскость.

Говоря о первом и втором этапах создания систем МП, мы особое внимание уделили конкретным лингвистическим коллективам. Характеризуя современный период, попытаемся определить основные теоретические и технические проблемы, связанные с выбором стратегии МП и с эффективностью использования практических систем МП. Кроме того, обсуждение этих основных проблем МП естественно включает в себя рассмотрение перспектив дальнейшего развития систем МП, как имеющих самостоятельную практическую ценность, так и входящих в многофункциональные системы автоматической переработки текстов (АПТ).

Первая проблема сводится к обсуждению трех возможных теоретических стратегий:

— прямого бинарного МП;
— перевода с трансфером;
— перевода с ЯП,

а также особенностей реализующих их грамматических средств (машинных, грамматических и собственно парсеров).

Прямой бинарный перевод, который реализован в «старых» системах МП и по-прежнему является основной стратегией многих практических систем МП, ориентирован на конкретную языковую пару (системы GAT, SYSTRAN и их дериваты, система АРМАС, разработанная в ВЦП). В процессе бинарного перевода осуществляются элементарные трансформации предложения на основе установления границ сегментов и контекстного снятия омонимии.

Опыт построения и эксплуатации давно построенных систем МП, например, системы Джорджтаунского университета GAT и ее вариантов, показывает, что эффективность этих систем обеспечивается путем использования мощных словарных средств, накопленных за десятилетия их функционирования, и специальных средств редактирования. Однако прямое бинарное сопоставление, осуществляемое на основе отдельных лек-

сико-морфологических и семантико-синтаксических подструктур, жестко ориентирует лингвистическое и программное обеспечение системы не только на особенности конкрентной пары языков, но и на специфику подъязыков. Замена одного из членов такой пары немедленно требует перестройки лингвистического, а иногда и программного обеспечения.

Трансформационный МП (перевод с трансфером) также опробован во второй период развития МП. Применение этой стратегии, предусматривающей использование трансфера, основывается на теоретическом предположении о невозможности создания записи для прямой связи текстов разных языков. Эта связь может быть установлена только путем последовательных преобразований на различных уровнях репрезентации: лексическом, морфологическом, уровне синтаксических структур (в терминах конфигурационного и реляционного синтаксиса) и на уровне синтаксического описания. Использование трансфера предполагает последовательный переход от деревьев зависимостей, описывающих структуру входного предложения на различных уровнях анализа, к глубинной ролевой структуре, используемой в качестве базового грамматического представления, и далее к выходным деревьям зависимостей. При этом использование процедуры преобразования именно для локальных деревьев зависимостей позволяет получать МП, не моделируя полное понимание предложения. Решение этой задачи не обязательно при МП, так как глубинные семантико-синтаксические интерпретации могут быть оставлены человеку-пользователю, т. е. специалисту, работающему с переводом, выданным ЭВМ. Такой подход к использованию трансфера характерен не только для большинства работающих зарубежных коллективов, но и для ряда отечественных систем МП.

В этих системах выделяют три основные стадии перевода: анализ, трансфер и синтез. Каждая из стадий реализуется как последовательность субпроцессов. Так, в уже упоминаемой системе SUSY анализ включает в себя 8 субпроцессов, результатом которых является построение дерева входного предложения, трансфер — 8 субпроцессов, результатом которых является дерево выходного предложения. Это дерево на этапе синтеза в результате взаимодействия трех субпроцессов преобразуется в предложение на ЯП. Аналогичный подход используется и в других системах с трансфером (CADA, GETA, TAUM-METEO, японский проект, проект EUROTRA, в нашей стране — система СИЛОД (MULTIS), разработанная в РГПУ им. А. И. Герцена, системы ФРАП, АРАП, ЭТАП-2, STYLUS и др.).

Предложенный именно в третий период развития МП проект EUROTRA основан в 1982 г. при Европейском Экономическом Сообществе особой комиссией (КЭС). Директор проекта С. Першке. Языки системы: английский, французский, немецкий, итальянский, испанский, португальский, голландский, датский,

новогреческий. Этот проект предусматривал создание системы МП для девяти официальных языков стран ЕЭС, рассчитанной на перевод 72 пар языков. Для унификации строго уровневого анализа выработано единое представление лингвистических данных и единые правила их обработки.

В результате анализа текста на любом из переводимых языков осуществляется семантическая интерпретация синтаксических отношений, так называемое интерфейсное представление. Текст представлен конструкциями семантических зависимостей (типа глубинных падежей). Интересно отметить, что интерфейсная структура одного языка не совпадает с таковой любого другого, но зато «разрешает» применение трансфера. Таким образом, каждому специалисту в области своего языка необходимо построить кроме анализа и синтеза 8 процедур трансфера.

Целью проекта являлось создание предпромышленного образца системы МП. Особенностью трансфера, реализуемого в системе EUROTRA, является пошаговость, при этом на каждом шаге трансфера результат «очищается» от неверных преобразований, которые были получены на предыдущем шаге.

К системам, использующим трансфер, относится система ЭТАП-2 (Апресян и др., 1989). Эта система англо-русского МП, идеологически являющаяся развитием системы ИНФ-ЭЛ, обладает особым словарным обеспечением, включающим в себя комбинаторные автоматические словари (АС), не ориентированные на конкретную предметную область. Возможность такой «независимости» связывается авторами с полнотой разработанных лингвистических моделей.

Все правила трансфера в системе делятся на общие, частные (трафаретные) и словарные. В системе выделяются 45 элементарных предикатов и 5 синтаксических отношений, фильтрующие механизмы и правила предпочтения позволяют отдать предпочтение одним синтаксическим гипотезам перед другими.

Использование трансфера всегда предполагает последовательный переход от поддеревьев, описывающих структуру входного предложения на различных уровнях анализа, к глубинной ролевой структуре, используемой в качестве базового грамматического представления, и далее к выходным поддеревьям. При этом использование процедуры преобразования именно для локальных поддеревьев позволяет получать МП, не моделируя понимания предложения. Задача моделирования понимания не входит в круг проблем МП, так как углубленные интерпретации могут быть оставлены человеку — специалисту в конкретной предметной области.

Такой подход к реализации систем МП с трансфером осуществлен и в системе СИЛОД, также возникшей во второй период развития МП (в современном варианте — многоязычная система SILOD-MULTIS, реализованная на персональных компь-

ютерах). Весь процесс МП моделируется в системе SILOD-MULTIS как композиция лексико-морфологического и семантико-синтаксического субпроцессов. В ходе лексико-морфологического анализа осуществляется идентификация текстовых и словарных единиц и извлечение словарной информации из блоков автоматического словаря. Результатом такого анализа является табличное представление текста, в котором каждой словоформе поставлено в соответствие ее лексико-синтаксическое описание, определенное с точностью до конверсионной и/или падежной омонимии, функциональные, ролевые и семантические характеристики, а также информация о переводных эквивалентах в канонической форме. Это представление, как и представление текста любого уровня анализа, может быть использовано различным образом (МП, аннотирование, реферирование и др.). Если говорить о МП, то его результатом на этом этапе анализа является лексико-морфологический и фразеологический подстрочник.

В ходе семантико-синтаксической переработки осуществляется межъязыковое преобразование структур, при котором используется вся информация, полученная на этапе лексико-морфологического анализа, а затем подключаются грамматические и семантические блоки. Этот трансфер моделируется как комплекс вертикально сопряженных подсистем, иерархия компонентов которого извлекается из текста.

Семантико-синтаксический анализ осуществляется последовательно на разных уровнях структуры предложения: уровне групп, уровне функциональных сегментов и, наконец, уровне предложения (Беляева, 1986; Апресян и др., 1989, Соколова, 1990). На уровне групп устанавливаются границы групп и связи ЛЕ. На следующем уровне выделенные последовательности объединяются в функциональные сегменты, соответствующие главным членам предложения. В выделенных сегментах осуществляется преобразование на уровне модификации табличного представления текста. Результатом, полученным на уровне функциональных сегментов (групп), является МП с ограниченной грамматикой. Полный же семантико-синтаксический анализ достигается на уровне предложения.

Синтаксическое описание является определяющим при осуществлении процедур анализа и трансфера, так как именно на его основе происходит распознавание возможного фрейма предложения. Ключевую роль в таком распознавании играет лексико-синтаксическое описание глагола, содержащее информацию о наборе актантов, возможной модели управления и т. д.

В соответствии с этой методикой SILOD-MULTIS реализован как комплекс автоматических словарей и грамматик. В базовой конфигурации системы предусмотрены возможности работы с индоевропейскими и иноструктурными языками на основе универсальной структуры входных словников и единой

структуры выходного русского словаря для всех входных языков.

Использование трансфера в реально действующих системах МП подтверждает целесообразность выбора именно этой стратегии. В то же время уровень развития систем искусственного интеллекта и новые возможности вычислительной техники определили новое пробуждение интереса к третьей стратегии МП.

Стратегия перевода с использованием ЯП отражает влияние систем МП и их идеологии при создании баз знаний на прагматику МП. ЯП, применявшиеся на первом этапе развития МП, значительно опередили свое время, и хотя сегодня мы можем говорить только об экспериментальной реализации этой идеи (система CMU университета Карнеги, DLT, разрабатываемая в Утрехте, JETR — в Японии, SAM и TRANSLATOR в США), само ее возрождение свидетельствует о новом этапе МП. Увязка систем МП, использующих ЯП, и систем искусственного интеллекта определяется необходимостью использовать в них сложные базы знаний, с помощью которых должно быть распознано значение предложения или содержание целого текста и автоматически сгенерирован его пересказ на выходном языке.

Моделирование концептуального уровня понимания, являющееся промежуточным уровнем систем искусственного интеллекта, основывается на нескольких базовых составляющих:

— языки представления знаний, требующие организации словарей в виде статей, включающих в себя узлы-понятия и типы связей (свойства);

— системные грамматики, осуществляющие преобразование по принципу признак — значение.

Основой ЯП на современном этапе являются языки представления знаний, при этом целью анализа является представление значения предложения, получаемого на основе входного предложения, дополненного и размеченного с помощью информации из базы знаний и в ее терминах. Процессу порождения перевода соответствует преобразование этой структуры в выходной язык. Естественно, результатом такого подхода является не собственно перевод, а, скорее, пересказ исходного текста или предложения.

Язык представления знаний, используемый в подобных системах, и является ЯП, осуществляющим семантическую «разметку» входной фразы относительно семантической сети и генерацию выходной фразы из этого внутреннего представления Такая генерация осуществляется по образцам (шаблонам, фреймам и т. п.). При всей общетеоретической привлекательности этого подхода он остается экспериментальным до тех пор, пока не будут созданы лингвистические основы и программные средства для построения мощных семантических сетей на основе автоматизированных лексиконов.

Широкая полевая эксплуатация систем МП позволяет перей-

ти от чисто умозрительной оценки качества перевода, получаемого в экспериментальном режиме, к классификации практических систем МП, включаемых в рабочее место переводчика, по степени их автоматизированности, особенностям редактирования и целевому назначению (Machine Translation, 1987, Апресян и др., 1989).

Рассматривая особенности систем МП на современном этапе развития, приходится отказываться от попыток классифицировать их по поколениям. Дело в том, что подобные классификации были созданы по аналогии с вычислительной техникой, в которой признак «поколения» непосредственно связан с элементной базой ЭВМ, действительно полностью заменяемой при переходе к следующему классу ЭВМ. В случае систем МП такой замены не происходит, так как ряд систем, традиционно относящихся к первому поколению, благополучно функционирует до сих пор (например, система GAT), в то время как некоторые системы следующих поколений так и не преодолели в экспериментальной проверке барьера отторжения формализации естественным языком.

Опыт разработки и использования различных современных систем МП показывает, что их анализ и сравнение целесообразно только по функциональным критериям, к которым относятся следующие.

1. Прагматика системы. Создание системы как промышленной или экспериментальной определяет ее основные характеристики и выбор методологии анализа. Этот критерий однозначно определяет требования к объемам словарной базы: не менее 20 000 словарных статей на одну пару языков в промышленной системе — ср. системы Bravice-Weidner LOGOS, METAL, GAT, SYSTRAN и т. д. Единственным исключением из этого правила является система TAUM-METEO, словарь которой включает в себя только 1500 словарных статей — это связано с очень жестким ограничением тематики входных текстов. В экспериментальных системах объем автоматических словарей «снизу» не ограничен, системы могут быть реализованы и при словарях в 200 словарных статей.

Цель создания промышленных МП ясна — это получение МП потоков документов, организованное как регулярный процесс. Промышленные системы МП должны отвечать следующим требованиям.

— Устойчивость работы системы. Система МП должна давать результат, который специалист может использовать даже в случае дефектов исходного перерабатываемого материала и неполноты словарного обеспечения. При этом «грубые» переводы могут быть использованы как исходный материал для получения качественного ручного перевода (ср. способ использования системы SYSTRAN).

— Тиражируемость системы. Система должна иметь доста-

точно простые программные и лингвистические средства для расширения областей ее применения. Кроме того, в системе должны выделяться универсальные модели, использование которых облегчает переход к новым предметным областям и языкам.

— Адаптируемость системы. Система МП должна иметь средства настройки на потребности конкретных пользователей и особенности обрабатываемых материалов.

— Оптимальность временных параметров. Скорость перевода текстов в системе МП должна соответствовать либо объему поступающей в единицу времени информации, либо нормам работы редакторов, переводчиков или специалистов, являющихся пользователями системы.

— Комфорт пользователя. Сервисные средства системы должны обеспечивать удобство работы пользователя как в пакетном, так и диалоговом режиме.

Целью создания экспериментальных систем является исследование возможных путей развития МП, моделирование речемыслительной деятельности человека. Именно в этом плане согласуются задачи МП и искусственного интеллекта. Экспериментальные системы и их разработчики выступают в роли первопроходцев, а получаемые ими результаты могут использоваться в промышленных системах. Так, например, на основе результатов, полученных при создании экспериментальной системы Саарбрюкенского университета SUSY, имеющей чисто исследовательские цели, создается промышленная система MARIS и разработано автоматизированное рабочее место переводчика SUSANNAH.

2. Степень автоматизации системы. Практически все системы МП требуют вмешательства человека, однако применение этого критерия позволяет классифицировать системы по отношению к двум полярным точкам: от МП с редактированием до систем в помощь к человеку-переводчику (автоматизированное рабочее место переводчика — ср. систему СПАС, разработанную во Всесоюзном центре переводов).

Естественно, основой различия этих систем является стратегия вовлечения человека — редактирование полученного системой результата или использование словарных мощностей системы при ручном переводе. Соответственно может различаться этап редактирования и его объем. Так, в промышленных системах МП наиболее широко используется постредактирование, охватывающее, как правило, 20% переведенного материала (ср. системы ALPS, ENGSPAN, LOGOS, METAL, SPANAM), особенно это характерно для систем, предусматривающих пакетную обработку массивов текстов. Предредактирование используется только в одной промышленной системе МП — TAUM METEO, которая ориентирована на перевод текстов канадского бюро погоды с английского языка на

французский. В этой системе, в свою очередь, не используется постредактирование. Интерредактирование используется в настоящее время только в экспериментальных системах.

Повсеместное обращение к построению практических систем и отход от «бумажных» алгоритмов выдвигает на первый план наравне с вопросами стратегии проблемы повышения эффективности использования результатов МП пользователями — специалистами и переводчиками.

Ясное понимание того факта, что результат МП является чаще всего первопродуктом, требующим последующего (или предварительного) редактирования, привело разработчиков практических систем МП к необходимости создания специального сервиса и, более того, к учету уровня этого сервиса при оценке эффективности системы МП в целом. Такой сервис включает в себя, как правило, средства редактирования, средства интерактивного введения словарных систем — автоматизированных лексиконов и т. п.

Новый виток информационного взрыва определяет необходимость широкого использования систем МП и, следовательно, учета перечисленных факторов.

Литература

Автоматический перевод. 1949—1963 гг.: Критико-библиографический справочник / Сост. И. А. Мельчук, Р. Д. Равич. М., 1967, 516 с.

Автоматический перевод: Сб. статей / Под ред. и предисл. О. С. Кулагиной, И. А. Мельчука. М., 1971. 357 с.

Андреев Н. Д. Машинный перевод и проблема языка-посредника // Вопр. языкозн. 1957. № 5. С. 117—121.

Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л. и др. Лингвистическое обеспечение системы «ЭТАП-2». М., 1989. 295 с.

Бар-Хиллел И. Будущее машинного перевода // Науч. докл. высшей школы. Филол. науки. 1962. № 4. С. 203—206.

Белопольская А. Р., Коврыгин А. Б. Исходный этап независимого анализа немецкого простого распространенного предложения // Материалы по математической лингвистике и машинному переводу. Вып. 2. Л., 1963. С. 150—160.

Беляева Л. Н. Применение ЭВМ в лингвистических исследованиях и лингводидактике. Л., 1986. 84 с.

Воронин В. А. Система независимого конфигурационного анализа китайского текста при машинном переводе // Машинный перевод. М., 1961. С. 135—172.

Дрейзин Ф. А., Рашитов Р. С. Принцип синтаксического анализа татарской фразы // Машинный перевод. М., 1961. С. 295—303.

Кремнева Н. Д., Морева В. Ю., Ульянова Г. С. Принципы построения грамматики русского синтаксического анализа с учетом предпочтения и порядка замыкания связей // Лингвистические проблемы функционального моделирования речевой деятельности. Вып. 3. Л., 1976. С. 139—144.

Кулагина О. С. Исследования по машинному переводу. М., 1979. 319 с.

Кулагина О. С., Мельчук И. А. Машинный перевод с французского языка на русский // Вопр. языкозн. 1956. № 5. С. 111—121.

Лейкина Б. М., Никитина Т. Н., Откупщикова М. И. и др. Система автоматического перевода, разрабатываемая в группе математиче-

ской лингвистики ВЦ ЛГУ // Науч.-тех. информация. 1966. № 1. С. 40—50.

Леонтьева Н. Н. Система французско-русского перевода (ФРАП): лингвистические решения, состав, реализация // Машинный перевод и прикладная лингвистика: Проблемы создания системы автоматического перевода. Вып. 271. М., 1987. С. 6—26.

Марчук Ю. Н. 1) Проблемы машинного перевода. М., 1983; 2) Методы моделирования перевода. М., 1985. 199 с.

Машинный перевод: Сб. статей. М., 1957. 314 с.

Мельчук И. А. 1) Работы по машинному переводу в СССР // Вестн. АН СССР, 1959. № 2. С. 43—47; 2) Опыт теории лингвистических моделей «смысл — текст». М., 1974. 314 с.

Молошная Т. Н. Алгоритм машинного перевода с английского языка на русский // Проблемы кибернетики. Вып. 3. М., 1960. С. 209—272.

Откупщикова М. И. Синтаксис связного текста. Л., 1982. 103 с.

Панов Д. Ю. Автоматический перевод. 2-е изд. М., 1958. 72 с.

Переводная машина в Таллине в 1924 г. // Сообщения по машинному переводу. Сб. 1. Таллин, 1962. С. 95—96.

Переводная машина П. П. Троянского: Сб. матер. о переводной машине для перевода с одного языка на другие, предложенной П. П. Троянским в 1933 г. / Отв. ред. Д. Ю. Панов. М., 1959. 52 с.

Пиотровский Р. Г. Инженерная лингвистика и теория языка. Л., 1979. 112 с.

Синтаксический компонент в системах машинного перевода // ВЦП. Сер. 2. Машинный перевод и автоматизация информационных процессов: Обзорная информация. Вып. 5. М., 1981. 130 с.

Шаляпина З. М. Англо-русский многоаспектный автоматический словарь (АРМАС) // Машинный перевод и прикладная лингвистика. Вып. 17. М., 1974. С. 7—67.

Шумилина А. Л. Вопросы анализа личных местоимений третьего лица // Лингвистические исследования по машинному переводу. Вып. 2. М., 1961. С. 132—140.

Цейтин Г. С., Откупщикова М. И., Лейкина Б. М. и др. Система анализа текста с процедурным представлением словарной информации // Актуальные вопросы практической реализации систем автоматического перевода. Ч. 1. М., 1982. С. 205—231.

Чижаковский В. А., Беляева Л. Н. Тезаурус в системах автоматической переработки текста. Кишинев, 1983. 163 с.

Hutchins W. J. Translation: Past, Present, Future. Harwood, 1986.

Machine Translation: Theoretical and Methodological Issues // Ed. by S. Nirenburg. New York, 1987. 350 p.

РЕФЕРИРОВАНИЕ

СВЕРТЫВАНИЕ ИНФОРМАЦИИ

Непрерывно растущий объем научной и научно-технической информации не позволяет ее потребителям отыскивать необходимые им материалы методом поиска в массивах первичных документов, поскольку такой поиск потребовал бы слишком много времени, труда и в конечном итоге оказался бы неэффективным, так как не обеспечил бы нужной полноты. Именно поэтому поиск ведется по массивам вторичных документов. К числу вторичных документов традиционно относят библиографическое описание, аннотацию, реферат и научный перевод. Вторичный документ представляет собой результат свертывания информации первичного документа. Под свертыванием принято понимать сжатие, или компрессию, текста первичного документа при его переработке в текст вторичного документа. Это свертывание максимально в библиографическом описании, потому что от текста там остается практически только заголовок, и минимально в научном переводе, потому что перевод с точки зрения объема и содержания информации должен быть эквивалентен подлиннику, если только перевод не производится с сокращениями, что стало в настоящее время широко распространенной практикой; аннотации и рефераты занимают промежуточное место.

Аннотация и реферат различаются выполняемыми ими функциями. Функция аннотации — всецело указательная. Аннотация только указывает на характер содержащейся в документе информации, не раскрывая ее и в редких случаях характеризуя. Реферат выполняет как указательную, так и информативную функцию, поскольку он с большей или меньшей степенью подробности раскрывает содержание первичного документа. Различию между аннотацией и рефератом соответствует в англоязычных странах различие между указательным рефератом (indicative abstract) и информативным рефератом

(informative abstract). Аннотация и реферат различаются также объемом. Чисто указательная функция аннотации требует значительно меньшего объема, реферат же для выполнения своей функции нуждается в гораздо большем объеме. Особенно это касается рефератов по гуманитарным наукам. Однако в любом случае объем реферата не может быть определен вполне точно, поскольку он зависит от многих факторов, в число которых входит характер первичного документа (статья, монография и т. д.), его доступность (труднодоступные документы реферируются обычно подробнее), язык первичного документа (документы на редких языках реферируются детальнее), наука или предметная область (естественная — гуманитарная) и т. д.

ИНТЕЛЛЕКТУАЛЬНОЕ РЕФЕРИРОВАНИЕ

Под интеллектуальным реферированием понимают реферирование первичного документа, производимое человеком, обычно профессиональным референтом. В настоящее время такое реферирование играет исключительно важную роль при ручном поиске информации, поскольку составляемые по такой методике рефераты образуют фундамент подавляющего большинства реферативных журналов.

Термин «интеллектуальное» отражает тот факт, что реферирование осуществляется именно человеком. Таким образом, интеллектуальное реферирование противопоставлено автоматическому реферированию, производимому с помощью ЭВМ. В последние годы роль автоматического реферирования непрерывно возрастает.

При рассмотрении реферирования важно проводить различие между авторским рефератом (summary) и рефератом, составляемым профессиональным референтом (abstract). Первый составляется самим автором первичного документа и отражает его понимание наиболее важных аспектов этого документа. Авторский реферат обычно предпосылается первичному документу (например, статье или монографии), однако составляется не всегда и далеко не всеми авторами.

Функция авторского реферата заключается только в информировании читателя о содержании следующего за ним текста. При этом данное информирование не очень сильно отличается от того указательного информирования, которое достигается аннотациями. Никаких поисковых функций авторский реферат не выполняет. Между тем обычный реферат, будучи помещенным в реферативный журнал, выполняет еще и поисковую функцию, поскольку при ручном поиске критерий релевантности документа обычно определяется относительно содержания реферата.

Реферируя те или иные документы (например, научные ста-

тьи), референты могут не только обращаться к тексту первичного документа, но и использовать текст авторского реферата, особенно если содержание его совпадает с представлением референта об основном содержании соответствующего первичного документа. Не следует, однако, забывать, что авторские рефераты, особенно в журналах промышленно развитых капиталистических государств, имеют нередко и явно выраженный рекламный характер, что не всегда позволяет по авторскому реферату судить об основном содержании документа.

По объему авторские рефераты обычно значительно короче рефератов, составляемых референтами. Конкретные результаты исследования или разработки в них также, как правило, не приводятся.

РЕФЕРАТИВНЫЕ ЖУРНАЛЫ

Библиографические описания, отражающие какую-нибудь предметную область или проблему, могут быть упорядочены по тематике или по фамилиям авторов, что дает возможность оформить их в виде книги, обычно называемой библиографическим указателем литературы (БУЛ). Каждое или некоторые библиографические описания могут быть снабжены аннотацией. В результате будет получен аннотированный БУЛ.

Библиографические описания, относящиеся к различным областям науки, могут наноситься на каталожные карточки. Эти карточки, будучи упорядоченными в каталожных ящиках по алфавиту фамилий авторов или по отраслям знания, дают возможность формировать авторские или систематические каталоги, которые наравне с БУЛами можно рассматривать как ручные информационно-поисковые системы (ИПС).

Сходным образом систематизируются и рефераты в реферативных журналах (РЖ), причем основой такой систематизации является принцип деления универсума знания на отрасли и предметные области с любой необходимой степенью детализации. РЖ также являются элементарными ИПС ручного типа, поскольку они облегчают отыскание необходимых сведений.

В настоящее время в мире используется большое количество РЖ, различающихся по тематике и структуре. В нашей стране основными видами РЖ являются РЖ ВИНИТИ и РЖ ИНИОН. Разумеется, существуют и другие РЖ, например, РЖ ВНИИМИ, РЖ ВИНТИСХ и др.

РЖ ВИНИТИ отражают документы по естественным и точным наукам, РЖ ИНИОН — по наукам гуманитарного профиля. Данные РЖ внешне различаются форматом, а внутренне — структурой и объемом рефератов. Рефераты РЖ ВИНИТИ значительно короче рефератов РЖ ИНИОН и имеют более каноническую форму, в соответствии с которой в начальной части

реферата указываются описанные в первичном документе проблемы, а в заключительной перечисляются полученные результаты. Реферат РЖ ИНИОН в несколько раз (до 10 и даже более раз) превышает по объему реферат РЖ ВИНИТИ и по структуре практически ничем не отличается от краткой статьи. Это и естественно, поскольку гуманитарную проблематику трудно свести к нескольким пунктам или к кратким выводам. Здесь важен и характер аргументации.

Реферат любого документа в РЖ дается не изолированно, а совместно с библиографическим описанием, включающим в себя фамилию автора, заголовок и выходные данные. Это дает возможность по-разному упорядочивать текст библиографического описания и текст реферата. Наиболее распространенными являются следующие три типа рефератов:

1) обычный реферат (англ. conventional abstract) с последовательностью: заголовок — библиографическое описание — текст реферата;

2) реферат с тематическим предложением на первом месте (англ. topical-sentence-first abstract или findings-oriented abstract). Порядок информации следующий: тематическое предложение — текст реферата — библиографическое описание. При этом под тематическим предложением понимается наиболее важное в информационном отношении предложение, извлеченное из текста реферата;

3) реферат с заголовком на первом месте (англ. title-first abstract), имеющий последовательность: заголовок — тематическое предложение — текст реферата — библиографическое описание (без повторения заголовка).

В различных РЖ используется тот или иной вид рефератов. Например, в РЖ "Chemical Abstracts" используется реферат типа 2, в РЖ ВИНИТИ — реферат типа 1.

В литературе по реферированию неоднократно обсуждался вопрос об оптимальном варианте интеллектуального реферата, т. е. о том, какие компоненты содержания первичного документа и в какой последовательности надлежит отражать в реферате. Многочисленные эксперименты, однако, показали, что оптимального варианта реферата для некоторого заданного текста первичного документа не существует. В принципе, для одного и того же текста первичного документа может быть составлено достаточно большое число различных рефератов, которые будут признаны вполне удовлетворительными со стороны специалистов. Поэтому данная проблема в последнее время получила несколько иную интерпретацию, а именно: вместо поиска идеального варианта реферата допускается составление нескольких рефератов по одному тексту, таких, что каждый из них рассчитан на потребности специалиста определенного профиля. Такой подход обусловлен тем, что специалисты с разными интересами обращают внимание на разные аспекты

содержания одного и того же первичного документа. Этот принцип широко пропагандируется и начинает в настоящее время воплощаться в жизнь также и в области автоматического реферирования.

АВТОМАТИЧЕСКОЕ РЕФЕРИРОВАНИЕ

Разработчики систем автоматического реферирования исходят из тезиса о том, что в тексте первичного документа существуют более и менее информативные фрагменты. При этом основная задача состоит в том, чтобы выявить наиболее информативные фрагменты, которые и могут лечь в основу автоматического реферата. В соответствии с существующими представлениями, максимум информации концентрируется: 1) в одной точке текста, 2) в нескольких так называемых фокальных, или опорных, точках текста, 3) не имеет строго определенной локализации.

Наиболее распространенным является взгляд о том, что носителями максимума информации являются опорные точки. Такие точки находятся обычно в начале или конце текста, а конкретно в таких частях текста, к которым относятся заглавие, введение, выводы, заключение, рекомендации, причем самыми главными из них считаются выводы и заключение. Существует, однако, и иная точка зрения, в соответствии с которой максимально информативные компоненты распределены по всему тексту более или менее равномерно.

При любой точке зрения на локализацию информативных компонентов текста необходима тщательно разработанная методика их выявления. В настоящее время известно большое число методик, которые базируются на различных признаках, присущих текстам и их частям. Все эти методики исходят из принципа фрагментирования, в соответствии с которым текст вначале разбивается на отрезки, а затем внутри них отыскиваются наиболее информативные предложения, словосочетания и слова. Существующие и широко используемые методы целесообразно рассмотреть более подробно.

Статистико-дистрибутивные методы. Статистико-дистрибутивный метод представляет собой единство дистрибуции как совокупности всех окружений рассматриваемой языковой единицы и статистики как количественных данных об этих окружениях. Статистический критерий обеспечивает объективность и алгоритмируемость статистико-дистрибутивных процедур, а следовательно, и возможность автоматизации этих методик.

Статистико-дистрибутивные методики нашли широкое применение в дешифровке, статистико-комбинаторном моделировании, в автоматическом индексировании и реферировании, а также в исследовании семантики.

В общей статистико-дестрибутивной методике, использу-

мой в целях автоматического реферирования, выделяется несколько частных методов. К их числу относится прежде всего метод Х. П. Луна, в соответствии с которым наиболее информативными считаются предложения, содержащие скопления значимых для данного документа слов, расположенных достаточно близко друг к другу. При этом позиционные ограничения не накладываются: информативные предложения могут отбираться из любого места первичного документа. Скоплением считается любая цепочка словоформ предложения, крайние слова которой являются для данного текста значимыми, причем между ближайшими значимыми словами цепочки находится не более пяти незначимых слов. Хотя метод Луна не учитывает смысловых связей между словами, его преимущество заключается в возможности его полной автоматизации.

Существует ряд модификаций метода Луна, имеющих целью его оптимизацию. Например, по методу В. А. Освальда информативность предложения оценивается не только наличием скоплений значимых, т. е. ключевых, слов, но и количеством таких скоплений. Согласно методу Л. Эрла информативными считаются предложения с числом скоплений не менее трех. По методу ACSI-Matic в скоплениях учитываются ключевые слова с частотой, превышающей среднюю частоту в документе. Л. Дойл и М. Квиллиан предложили методику учета совместной встречаемости терминов, или ключевых слов, в предложениях и вычисления коэффициента подобия с помощью особой матрицы взаимосвязанных элементов.

Принцип совместной встречаемости использовался также и другими исследователями, например А. Я. Шайкевичем, однако не в целях автоматизации реферирования, а просто как метод исследования семантики.

Статистико-дистрибутивные методы позволяют простейшим образом выделить из текста те предложения, которые можно считать наиболее информативными и которые в силу этого можно включать в реферат. Для ограничения сферы поиска таких предложений статистико-дистрибутивные методы могут использоваться совместно с позиционными методами.

Недостаток статистико-дистрибутивных методов заключается в том, что между информативностью фрагмента и частотой входящих в него ключевых слов нет прямой и однозначной зависимости. Все оказывается значительно сложнее. Поэтому кажущаяся простота этого метода не всегда приводит к желаемым результатам. Кроме того, реферат, составленный только на основе статистико-дистрибутивной методики, не всегда отличается достаточной связностью и разительно отличается от интеллектуального реферата для того же текста. Поэтому статистический критерий по необходимости должен дополняться другими критериями.

Методы использования содержательных аспектов. Стати-

стико-дистрибутивные методы ничего не говорят о конкретном смысловом содержании выделяемых фрагментов текста. Поэтому об их информативности можно судить лишь косвенно. Однако содержание вычленяемых фрагментов можно существенно конкретизировать, если разработать процедуру, которая дала бы возможность помечать эти фрагменты смысловыми индикаторами. Такие индикаторы нередко называют содержательными аспектами, а основанный на них метод реферирования называют поаспектным. Среди разработчиков этого метода можно назвать В. И. Соловьева, Н. А. Пащенко, С. А. Карасева, Н. И. Гендину, Д. И. Блюменау, И. С. Добронравова, Д. Г. Лахути, В. П. Леонова, Е. Б. Федорова и др.

Разные авторы предлагают различные наборы аспектов, однако чаще всего среди них фигурируют такие аспекты, как предмет исследования, цель и методы исследования, актуальность, специфика предлагаемого подхода или решения, преимущества, области применения, выводы, результаты, рекомендации и т. д. Иногда такие аспекты объединяются в семантические блоки. Каждый аспект обычно маркируется типовыми выражениями. Например, аспекту «цель исследования» чаще всего соответствует выражение «Цель настоящего исследования состоит в том, чтобы», аспекту «авторский подход» — выражение «В настоящей работе рассматривается проблема» и т. д.

Значение поаспектного метода реферирования достаточно очевидно. Он дает возможность включать в реферат те предложения и фрагменты, которые связаны с определенным содержанием и которые в силу этого могут рассматриваться как наиболее информативные. По наблюдениям многих исследователей, к числу наиболее информативных аспектов относятся аспекты «предлагаемый вариант решения» и «выводы». В число важных аспектов включается нередко и то, что выражено в заглавии.

Поаспектный метод дает возможность выделить из текста первичного документа своего рода зоны важности и приписать каждой такой зоне весовой коэффициент, ориентируясь на который, можно решать вопрос о том, включать фрагменты данной зоны в реферат или нет. Поаспектный метод можно использовать в комбинации со статистико-дистрибутивным.

Рефераты, полученные с помощью поаспектного метода, представляют собой ряд фрагментов текста, недостаточно связанных друг с другом, хотя и достаточно информативных для пользователя. Степень этой несвязности можно снизить использованием метода текстовых связей.

Методы текстовых связей. Информационная значимость предложения в тексте может определяться не только статистическими свойствами его компонентов и содержательными аспектами, но и характером и числом его связей с другими пред-

ложениями. На этом принципе построены модели автоматического реферирования, разработанные В. Е. Берзоном, а также С. М. Приходько и Э. Ф. Скороходько, И. П. Севбо и др.

В соответствии с одним из методов для отбора информационно наиболее значимых предложений задается некоторая пороговая величина, выражающая число связей. Регулируя эту величину, можно менять информационную насыщенность предложений и, следовательно, объем реферата. При этом характер самих предложений может не приниматься во внимание.

Другой, более сильный вариант этого метода заключается в том, что в тексте разграничиваются эксплицитные и имплицитные сверхфразовые единицы. При этом оказывается, что основное содержание несет обычно первое предложение эксплицитной сверхфразовой единицы. Это предложение, как правило, является автосемантическим. Информативность первого предложения эксплицитной сверхфразовой единицы возрастает при увеличении длины этой единицы. К числу информативных относятся также крайние предложения с многократными покрытиями, под которыми понимаются перекрывающие межфразовые связи.

Учет параметра межфразовых связей делает результирующий реферат более связным и целостным. Однако и такие рефераты существенно уступают по качеству интеллектуальным рефератам.

Комбинированные методы. Ввиду ограничений, внутренне присущих каждому отдельному методу выделения наиболее информативных фрагментов текста, вполне естественной представляется тенденция использовать комбинацию методов в расчете на получение более удовлетворительных результатов. Различного рода суждения на этот счет высказывались в работах В. П. Леонова, С. И. Гиндина, Г. В. Бондаренко, О. И. Яровенко, А. И. Новикова и Б. В. Якушина.

Почти все исследователи указывают на важность учета частоты слова в тексте или фрагменте. Не менее важной считается позиция предложения или иного фрагмента в тексте, а в ряде случаев и аспект текстовой связи между ними. Предлагается также принимать во внимание интервалы между ключевыми словами, ключевые слова, входящие в состав заглавий, тип текста. Значительно реже отмечается такой фактор, как морфологическая форма и синтаксическая функция слова или цепочки слов.

Ясно, что при выделении информационно значимых компонентов, которые должны войти в результирующий реферат, должен учитываться определенный порядок в использовании методов. Действительно, методы, ориентированные на лексику, например статистический, необходимы для того, чтобы выделить наиболее информативные предложения или их комплек-

сы. При этом такое выделение может корректироваться позиционными и аспектными методами. Однако необходимой текстовой связности реферата можно достичь учетом критерия межфразовых связей. Поэтому порядок применения методов реферирования может соответствовать естественному движению по схеме снизу вверх, а именно по схеме: лексика — предложение — позиционно или аспектно обусловленный комплекс предложений — межфразовые связи — реферат. Однако эта схема не единственная, и ей может быть противопоставлена противоположная схема движения сверху вниз, в соответствии с которой вначале осуществляется исходное фрагментирование всего текста первичного документа, затем в рамках выделенных фрагментов могут выделяться позиционно или аспектно характеризуемые предложения, после чего происходит погружение в лексический уровень. Оба направления можно использовать одновременно как две встречные методики.

ИНТЕЛЛЕКТУАЛЬНЫЙ РЕФЕРАТ И РЕФЕРАТЫ, СОСТАВЛЯЕМЫЕ С ПОМОЩЬЮ АВТОМАТИЧЕСКИХ ПРОЦЕДУР

Интеллектуальное реферирование имеет целью удовлетворение информационных или поисковых потребностей пользователей. Иначе говоря, с данными рефератами имеют дело прежде всего люди. Разумеется, такие рефераты могут использоваться и для других целей, например, для ручного или автоматического индексирования документов, для составления терминологических словарей и информационно-поисковых тезаурусов, для атрибуции текстов и т. д. Однако все эти цели для интеллектуальных рефератов являются побочными.

Что касается автоматического реферирования, то его цели во многих отношениях отличаются от целей интеллектуального реферирования. Разумеется, и такие рефераты могут быть ориентированы на потребности пользователя, в качестве которого выступает человек. Однако возможности машинных рефератов пока что несопоставимы с возможностями интелектуальных рефератов.

Машинные рефераты очень часто используются для целей индексирования, причем здесь вполне подходят рефераты, представляющие собой несвязную или малосвязную последовательность предложений или их фрагментов. Основное требование, предъявляемое к таким рефератам, — достаточное для индексирования документа количество ключевых слов. Такие рефераты обычно именуют термином квазиреферат.

Второй областью применения квазирефератов может быть атрибуция текстов, в соответствии с которой тексту должен быть приписан индекс предметной области, к которой данный

текст относится. Ясно, что высокая степень связности реферата и в данном случае не требуется.

Таким образом, квазирефераты — это своего рода лексикоцентрированные рефераты, тогда как интеллектуальные рефераты — это рефераты дискурсноцентрированного типа, т. е. рефераты, удовлетворяющие критериям связного текста. Это различие является принципиальным. Однако из этого вовсе не следует, будто принципиально невозможна разработка такой системы автоматических процедур, которая позволила бы существенно приблизить машинный реферат к параметрам интеллектуального реферата. Для этой цели необходимо не только использование комплекса разнообразных методик и процедур автоматического реферирования, но и их оптимизация. Более того, такой подход должен позволить подстройку под индивидуальные потребности пользователя, т. е. обеспечить необходимое прагматическое фокусирование.

Литература

Бондаренко Г. В. Об использовании формальных методов анализа структуры целого текста для автоматического индексирования и реферирования // Научно-техническая информация. Сер. 2. 1975. № 5.

Гендина Н. И. Выявление аспектов содержания документов с помощью формальных текстовых признаков // Научно-техническая информация. Сер. 2. 1979. № 4.

Гиндин С. И. 1) Методы автоматического фрагментирования текста, опирающиеся на характеристики внутреннего состава фрагмента // Семиотика и информатика. Вып. 9. М., 1977; 2) Позиционные методы автоматического фрагментирования текста, их теоретико-текстовые и психолингвистические предпосылки // Семиотика и информатика. Вып. 10. М., 1978.

Горькова В. И., Борохов Э. А. Реферат в системе научной коммуникации: Направления совершенствования лингвистических и структурных характеристик // Итоги науки и техники. Сер. Информатика. М., 1987.

Добрускина Э. М., Берзон В. Е. Синтаксические сверхфразовые связи и их инженерно-лингвистическое моделирование. Кишинев, 1986.

Карасев С. А. 1) Построение системы автоматического реферирования, включающей анализ и синтез текста (АНАСИН) // Научно-техническая информация. Сер. 2. 1978а. № 5; 2) Система автоматического реферирования: результаты машинного эксперимента // Научно-техническая информация. Сер. 2. 1978б. № 8.

Леонов В. П. О методах автоматического реферирования (США, 1958—1974) // Научно-техническая информация. Сер. 2. 1975. № 6.

Новиков А. И., Якушин Б. В. Алгоритм индексирования текстов взвешенными ключевыми словами по методу семантической фильтрации // Научно-техническая информация. Сер. 2. 1972. № 6. С. 15—20.

Пащенко Н. А., Кнорина Л. В., Молчанова Т. В. и др. Проблемы автоматизации индексирования и реферирования // Итоги науки и техники. Т. 7. М., 1983.

Севбо И. П. Структура связного текста и автоматизация реферирования. М., 1969.

Скороходько Э. Ф. Семантические связи в лексике и текстах // Вопросы информационной теории и практики (М.). 1974. № 23.

Шайкевич А. Я. Дистрибутивно-статистический анализ в семантике // Проблемы и методы семантических исследований. М., 1976.

ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ АВТОМАТИЗИРОВАННОГО ИЗДАТЕЛЬСКОГО ПРОЦЕССА

Автоматизация издательского процесса (АИП) не только позволяет переложить «на плечи» электронных вычислительных (ЭВМ) и фотонаборных (ФНМ) машин воспроизведение текста *авторского оригинала* в полиграфическом виде, но и разрешает автоматизировать целый ряд процессов его лингвистической обработки. Рассмотрим особенности необходимого для такой обработки лингвистического обеспечения (ЛО) по этапам издательской обработки текста. С этой целью предварительно перечислим эти этапы и дадим их кратную характеристику.

Этапы АИП. Автоматизированная технология выпуска изданий несколько изменила границы и специфику этапов традиционной технологии. С учетом сказанного в настоящее время можно выделить следующие этапы АИП:

1) авторский этап (состоит в записи автором текста своего произведения на машинный носитель информации);

2) редакционный этап, который включает в себя: а) корректуру и редактирование текста авторского оригинала; б) редакционно-издательскую деятельность (рецензирование, обработку корреспонденции, поиск информации в базах данных и др.);

3) этап конструирования (состоит в задании полиграфических параметров для каждого элемента текста: формата набора и полосы, кегля и гарнитуры шрифтов и т. д., — а также меток для выделения элементов текста);

4) этап полиграфического воспроизведения (состоит в отработке полиграфических параметров для каждого элемента текста с помощью ФНМ);

5) этап тиражирования, т. е. печатания текста в заданном количестве экземпляров (присутствует только при необходимости выпуска издания в традиционном «бумажном» виде);

6) этап распространения (состоит в передаче информации из банков изданий к читателям с помощью носителей информа-

ции или через электронные каналы связи, что в целом образует издательскую сеть).

На этапе 5 ЛО отсутствует полностью. На остальных этапах, кроме этапа 2, ЛО не является особенно мощным и разветвленным. Наиболее мощным оно является на этапе 2.

Поскольку этапы АИП хронологически могут быть совмещены (например, автор, набирая текст, сам осуществляет на первом этапе предварительное редактирование текста), то одни и те же функции ЛО могут встречаться на разных этапах АИП.

Рассмотрим особенности ЛО для каждого из перечисленных этапов в отдельности.

Авторский этап. На данном этапе авторы для набора текста используют различные текстовые процессоры, т. е. программы, позволяющие набирать и модифицировать текст. В такие текстовые процессоры часто встраивают следующие лингвистические функции:

1) поиск синонимов указанного автором слова и — при необходимости — автоматическая замена этого слова на синоним;

2) поиск слов, содержащих ошибки или отсутствующих в эталонном словаре орфографически правильных слов;

3) перенос слов при достижении заданного формата.

Первая функция используется только на данном этапе, а вторая и третья — и на последующих этапах, где они составляют основу этих этапов. Поэтому более детально эти функции будут рассмотрены ниже при описании соответствующих этапов.

На данном этапе автор может также использовать специальный формализованный язык меток для выделения в тексте его элементов. Если автор этих меток в тексте не проставляет, то на этапе конструирования их обязан присвоить элементам текста технический редактор.

Кроме того, на данном этапе можно автономно использовать толковые и энциклопедические словари, которые записаны на машинные носители информации и содержат программное обеспечение, позволяющее осуществлять быстрый доступ к нужной автору словарной статье.

Редакционный этап. На данном этапе АИП реализуют такие функции по обработке текста, как корректура и редактирование. Кроме того, параллельно с обработкой текста осуществляется редакционно-издательская деятельность. Перечисленные функции реализуются как в специально предназначенных для этих целей редакционно-издательских системах, так и в других типах систем, используемых в АИП, в которые встроены эти функции.

Тексты произведений как объект корректуры и редактирования в АИП состоят из следующих элементов:

— простые тексты (графемы, т. е. буквы, морфемы, слова,

словосочетания, предложения, сверхфразовые единства, блоки, т. е. разделы, главы и т. п., и дискурс, т. е. текст как целое);

— специальные тексты, например, рубрики, сокращения, наименования единиц измерений и т. п.;

— сложные тексты, т. е. формулы и таблицы;

— тексты, образующие аппарат произведений (титуальный лист, выходные сведения, указатели и т. п.);

— служебные тексты, т. е. метки элементов текста и команды полиграфического оформления.

Предметом обработки авторских текстов в АИП на данном этапе является два процесса: 1) приведение текстов в соответствие с оригиналом; 2) приведение текстов в соответствие с существующими нормами. В АИП первый из этих процессов принято называть корректурой, а второй — редактированием. В корректуре и редактировании можно выделить два типа операций: а) операции контроля, т. е. нахождения отклонений; б) операции реконструкции, т. е. исправления отклонений, найденных в процессе контроля.

Приведенное позволяет констатировать, что весь текст является объектом операций контроля, и только те его сегменты, в которых найдены отклонения, — объектом операций реконструкции. Среди отклонений следует различать (Партыко, 1989): а) искажения, т. е. отклонения от оригинала; б) ошибки, т. е. отклонения от норм; в) погрешности, т. е. отклонения от оригинала и норм одновременно.

Очевидно, что с точки зрения темы настоящего раздела нас интересуют не все те нормы, соблюдения которых следует контролировать в процессе редактирования (издательские, логические, технические и др.), а только лингвистические и психолингвистические. Среди этих групп норм можно выделить следующие их типы: а) списки (например, различные словари, перечни сокращений или перечни единиц измерений); б) шаблоны (например, шаблоны на расстановку знаков препинания в предложении или при перечислениях); в) структуры, или модели (например, структура издания или модель предложения); г) параметры (например, наличие на полосе не более четырех переносов подряд или наличие в аннотации не более 600 знаков).

Для определения качества текста, полученного от автора, прошедшего корректуру или редактирование, целесообразно использовать следующие критерии:

— для корректуры — степень искаженности текста копии:

$$E = e/M,$$

где e — число искажений в тексте, M — объем текста в авторских листах;

— для редактирования — степень ошибочности:

$$S = r/M,$$

где r — число ошибок в тексте.

Укажем также на некоторые технологические особенности машинной корректуры и машинного редактирования. Первой из таких особенностей является наличие трех этапов обработки текста:

— предкорректура (предредактирование), например присвоение элементам текста соответствующих меток;

— интеркорректура (интерредактирование), т. е. выполнение с помощью ЭВМ контроля и реконструкции;

— посткорректура (постредактирование), т. е. выполнение корректором или редактором тех операций, которые не могут выполнить технические средства.

Вторая технологическая особенность состоит в различной степени автоматизации корректуры и редактирования:

— компьютеризованная корректура и компьютеризованное редактирование (человек вручную осуществляет контроль и реконструкцию текста, записанного в памяти ЭВМ, через посредство текстового процессора);

— автоматизированная корректура и автоматизированное редактирование (контроль текста осуществляет ЭВМ, а реконструкцию — корректор или редактор);

— автоматическая корректура и автоматическое редактирование (контроль и реконструкция текста выполняет ЭВМ, человек вмешивается в нестандартных ситуациях).

Для определения эффективности методов контроля или реконструкции, а также корректуры или редактирования в целом следует использовать выражение

$$E = (O_1 - O_2)/O_1,$$

где E — эффективность обработки текста, O_1 — количество отклонений в тексте до обработки, O_2 — количество отклонений в тексте после обработки (для контроля при условии, что все найденные отклонения будут корректно исправлены).

Как свидетельствует практика и подтверждают научные исследования, при обработке потока текстов из них нельзя устранить ни всех искажений, ни всех ошибок. Поэтому более правильно будет говорить не о том, откорректированы или отредактированы тексты или нет, а о том, какова степень их обработанности (степень искаженности или степень ошибочности). Учитывая это положение, в литературе приводятся требования к степени искаженности для различных видов литературы.

Учитывая изложенное, опишем (Партыко, 1983) наиболее часто используемые методы машинной корректуры. Их можно подразделить на методы машинного контроля и методы машинной реконструкции. В свою очередь, методы машинного контроля можно подразделить на такие, которые базируются на использовании оригинала, и такие, которые базируются на использовании цепочек символов, соответствующих установленной норме.

Наиболее широкое применение среди методов, базирующихся на использовании оригинала, получил так называемый метод дублирования. Его суть состоит в том, что текст дважды записывают на машинный носитель информации, а после этого с помощью программы посимвольно сравнивают. Наличие в тексте в одной и той же позиции различных символов свидетельствует о том, что один из них является искаженным. Эффективность такого метода контроля значительно выше, чем у корректора-профессионала, и равна 0,96.

Среди методов контроля, базирующихся на использовании нормированных цепочек символов, наибольшее распространение получил так называемый словарный метод. Суть этого метода состоит в том, что орфографический словарь какого-то языка записывают на внешнее запоминающее устройство ЭВМ. Далее каждое слово корректируемого текста проверяют на наличие в эталонном словаре ЭВМ. Если слово текста в словаре найдено, то оно считается правильным, а если не найдено, то потенциально искаженным. Эффективность такого метода контроля ниже, чем у дублирования, и колеблется в пределах от 0,5 до 0,7.

Здесь может возникнуть вопрос: как можно с помощью норм (орфографического словаря) контролировать отклонения копии текста от оригинала? Объясняется такая возможность тем, что в большинстве случаев отклонения от оригинала одновременно являются и отклонениями от норм.

Относительно реконструкции необходимо отметить, что в настоящее время в издательской практике корректура практически всегда является автоматизированной, т. е. предполагает автоматическое нахождение искажений и их ручное исправление. Тем не менее существуют методы, позволяющие для формализованных текстов осуществлять реконструкцию с эффективностью от 0,8 до 0,9.

При осуществлении машинного редактирования количество контролируемых норм определяет «редактирующую» мощность систем редактирования. В значительной степени эта мощность зависит также и от того, насколько выделены элементы текста произведения с помощью меток, предназначенных для этой цели.

Лингвистические нормы позволяют проверять в основном простые тексты. Такой контроль базируется на словаре, снабженном для каждого слова различной информацией (морфологической, синтаксической, семантической, прагматической и др.). Наиболее простые из лингвистических методов позволяют, например, использовать только некоторое множество заранее отобранных по каким-то критериям слов или, наоборот, запрещает использовать некоторые слова. Более сложные методы могут осуществлять контроль правильности грамматической

связи между словами, расстановки знаков препинания, деления текста на абзацы, сокращений слов и словосочетаний и др.

Имеется достаточно широкий спектр психолингвистических норм (Мучник, 1985). Эти нормы всегда ориентированы на какую-то читательскую аудиторию и фиксируют максимально допустимую длину слов, предложений или сверхфразовых единств (эти характеристики часто называют еще читабельностью). Кроме того, для некоторых видов литературы устанавливают специфические нормы. Так, для учебников иностранных языков проверяют:

— новизну текста, т. е. отношение тех слов текста, которых учащийся не знает, к общему количеству слов в тексте (нормативное значение — от 0,015 до 0,050);

— адаптированность текста, т. е. отношение слов текста, которые учащемуся уже известны, к его общему словарному запасу (нормативное значение — от 0,237 до 0,253);

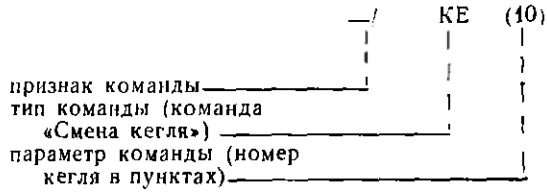
— напряжение изложения, т. е. отношение тех слов текста, которых учащийся не знает, к тем словам текста, которые учащемуся уже известны.

ЛО используют и для автоматизации редакционно-издательской деятельности. Для этих целей, в частности, разрабатывают и используют различные специальные информационно-поисковые языки, как, например, для обработки поступающей в издающую организацию корреспонденции.

Этап конструирования. На данном этапе для конструирования издания используют два типа языков: а) язык команд полиграфического оформления; б) язык меток для выделения элементов текста. Команды и метки могут состоять из 1, 2, 3 и более знаков.

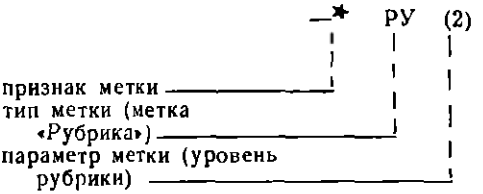
В настоящее время в нашей стране стандартов на такие формализованные языки нет, а поэтому в разных системах эти языки строятся на разных принципах. Количество меток и команд в системах наиболее часто находится в пределах от 20 до 100.

В большинстве случаев полиграфические команды имеют следующую структуру: а) признак команды; б) тип команды; в) параметр команды. В качестве примера приведем команду, состоящую из 8 знаков, одной из систем переработки текста:



Круглые скобки отделяют параметр от типа команды.

Метки элементов текста в основном имеют такую же структуру, как и полиграфические команды. Приведем пример такой метки:



В ряде систем язык меток элементов текста не используют, а вместо него перед каждым элементом текста ставят команду полиграфического оформления.

Команды и метки могут расставляться на любом из этапов, предшествующих этапу полиграфического воспроизведения текста.

Этап полиграфического воспроизведения. На данном этапе АИП из средств ЛО для выключки строк на заданный формат используют лишь одну функцию, реализуемую в виде программы переноса слов. Такие программы используют в автоматизированных системах переработки текстов, которые предназначены для отработки команд полиграфического оформления. Программы переноса слов, как отмечалось выше, могут использоваться и в других системах, которые функционируют на предыдущих этапах АИП (текстовых процессорах и редакционно-издательских системах), однако в этих системах к программам переноса предъявляют менее жесткие требования, чем на данном этапе.

Алгоритмы переноса слов можно разбить на два больших класса:

— словарные (в этом случае в памяти ЭВМ хранятся словари с эталонными переносами для 20—40 тыс. слов какого-то языка; такие словари могут предусматриваться для нескольких — чаще не более 10 — языков);

— несловарные (в этом случае переносы проставляются программой, в которой правила переноса либо реализованы в самом алгоритме программы, либо задаются в виде отдельного массива правил переноса).

Конечно, словарный метод переноса дает более высокие результаты, однако требует больших ресурсов (по времени и памяти). Несловарный метод, являясь более эффективным по времени и памяти, обеспечивает более низкое качество переносов.

Среди несловарных методов можно выделить следующие их подклассы:

— морфемные (ставят переносы в словах с учетом морфемного состава слов);

— безморфемные (ставят переносы на границах слогов или в зависимости от наличия в составе биграммы, т. е. двухбуквенного сочетания, гласных и согласных букв);

— комбинированные (ставят переносы с учетом префиксов, суффиков, а также слогов или биграмм).

Как показывают проведенные для несловарных методов исследования, количество неправильных переносов связано с количеством утерянных переносов обратной функциональной зависимостью. Так, если для одного из алгоритмов количество неправильных переносов равнялось 1,12%, то количество утерянных составило 48,41%. В противовес этому для другого алгоритма количество утерянных переносов составило 7,40%, а количество неправильных возросло до 11,70% (Будаар и др., 1985).

Этап распространения. На данном этапе АИП с помощью машинных носителей информации или каналов связи распространяют имеющиеся публикации либо в форме непериодических изданий, либо в виде периодических изданий (так называемых «электронных» журналов и газет). Такие «электронные» периодические издания имеются в целом ряде стран Западной Европы, Северной Америки и Японии (Мартин, 1987).

Для компоновки из отдельных публикаций периодических изданий на данном этапе АИП из средств ЛО используют в основном только различные рубрикаторы. На одном уровне иерархии в таких рубрикаторах наиболее часто допускается наличие (по техническим причинам) не более 10 рубрик.

Структура лингвистического обеспечения АИП. ЛО АИП может быть разделено на автономное и системное. Автономное обеспечение включает в себя толковый словарь, энциклопедические словари. Системное, в свою очередь, делится на блоки: опорная база данных, массив норм, программы.

Опорная база данных включает в себя: словари антонимов, синонимов, орфографический, словарь с морфологической, синтаксической, семантической и прагматической информацией (общеязыковой и отраслевой), словари эталонных переносов (для разных языков), списки морфем, единиц измерений, сокращений, персоналий, полиграфических команд, меток элементов текста, стандартных элементов библиографических описаний, специализированные информационно-поисковые языки, рубрикаторы.

Массив норм включает в себя массивы лингвистических норм отдельно по уровням: графем, морфем, слов, словосочетаний, предложений, сверхфразовых единств, блоков, дискурса; массив психолингвистических норм для читателей с высоким уровнем подготовки, с низким уровнем подготовки, массивы правил переноса (для разных языков).

Блок программ содержит программы для авторского этапа, для редакционного этапа, для этапа конструирования, для этапа полиграфического воспроизведения, для этапа распространения.

Каждый компонент структуры можно представить как отдельный объект, хотя некоторые компоненты могут быть совмещены (например, словарь с лингвистической информацией мо-

жет быть совмещен с орфографическим словарем и словарем синонимов). И для этого есть все необходимые технические предпосылки. Эти предпосылки состоят в следующем.

За последние годы на базе персональных ЭВМ и лазерных принтеров созданы так называемые настольные издательские системы (НИС). В этих НИС, которые имеют средства для работы и в режиме сети, можно реализовать практически все этапы АИП, кроме последнего. Поэтому возникает проблема создания такого компактного и одновременно мощного ЛО, которое позволяло бы, например, с помощью одного лингвистического словаря производить и поиск синонимов, и корректуру, и редактирование, и перенос слов.

В поисках путей создания такого компактного и мощного ЛО, на наш взгляд, и будет состоять ближайшая задача разработчиков лингвистического обеспечения АИП.

Литература

Будаар Я. И., Карская С. Д., Партыко З. В. Сравнительный анализ алгоритмов переносов слов русского языка // Научно-техническая информация. Сер. 2. 1985. № 12. С. 24—30.

Мартин Дж. Видеотекст и информационное обслуживание общества. М., 1987. 184 с.

Мучник В. С. Человек и текст: Основы культуры письменной речи. М., 1985. 256 с.

Партыко З. В. 1) Методы машинной корректуры и машинного редактирования. М., 1983. 41 с. (Изд. дело: Обзорн. информ. / НИЦ «Информпечать». Вып. 5); 2) Статистика ошибок при корректуре и редактировании текстов. М., 1989. 56 с. (Изд. дело: Обзорн. информ. / НИЦ «Информпечать». Вып. 3).

ЛИНГВИСТИКА В КАРТОГРАФИИ

Простейшие картографические рисунки были известны еще в первобытном обществе, задолго до зарождения письменности. Карта, по выражению известного ученого-географа Н. Н. Баранского, — особый язык, без которого не может обойтись ни один специалист, работающий на нашей планете и в околоземном пространстве. Инженер и исследователь, летчик и строитель, геолог и агроном, офицер и синоптик, ученый и государственный деятель — все обращаются к карте и находят ответы на свои вопросы.

Можно сказать, что карта — один из важнейших источников знаний, причем тексты на естественном, человеческом языке не обладают такой точностью и наглядностью, не позволяют оценить пространственные взаимосвязи объектов действительности.

Научное определение термина *карта,* содержащееся в стандарте картографических терминов, таково: построенное в картографической проекции, уменьшенное, обобщенное изображение поверхности Земли, поверхности другого небесного тела или внеземного пространства, показывающее расположенные на них объекты в определенной системе условных знаков.

Картографию понимают как науку о географических картах, методах и технологиях их создания и использования, карты рассматриваются в качестве способа отображения действительности и средства изучения реальных явлений.

Предмет картографии как науки и технологии вызывал много споров, особенно в 70—80-е годы: в картографию включались научные исследования и технические работы по созданию карт; карты рассматривались как произведения искусства; картография понималась как формальная наука о логике, методике и технике изготовления и использования карт; наконец, картография рассматривалась как часть информатики.

Трудно возразить против точного определения: предметом картографии является картографическая форма передачи информации. Такое понимание сближает картографию и лингвистику, картографию и семиотику. В самом деле, предмет лингвистики — естественный язык как средство и форма передачи информации; предмет семиотики как науки о знаковых системах — разнообразные системы знаков (например, знаки шахматных нотаций — средства записи шахматных партий, знаки уличного движения и т. д.), используемые для передачи информации.

Карта — образно-знаковая модель действительности, передающая пространственно-временную локализацию объектов действительности. Положение объекта всегда фиксировано в определенной пространственной системе координат, время существования объекта фиксировано временными рамками изображения и отмечается датами, содержащимися в зарамочном оформлении карты.

Информация, фиксированная и передаваемая картой, называется картографической информацией. Картографическая информация — это особый вид научно-технической информации, характеризующийся использованием двух знаковых систем для отображения предметных областей действительности: 1) естественного языка (ЕЯ), 2) особых типов семиотических систем — искусственных систем условных знаков (СУЗ).

Картографическая информация присутствует в любом картографическом документе вне зависимости от его содержания и объединяет три типа информации:

— смысловую (семантическую) информацию, передающую совокупность показанных на карте объектов и их характеристик при помощи СУЗ и терминов ЕЯ;

— метрическую информацию, передающую пространственное положение объектов относительно друг друга и земной поверхности при помощи принятой системы координат;

— собственно графическую информацию, представляющую собой сведения о конфигурации, цвете и размерах условных знаков карты.

В картографии широко используется термин *язык карты*, не имеющий, однако, единого значения. Под языком карты понимают: 1) искусственный язык графических символов — систем условных знаков, используемых для передачи информации (Салищев, 1976); 2) знаковую систему, состоящую из двух особых языковых слоев, или подъязыков, первый из которых состоит из системы картографических знаков — графических символов, передающих содержание карты, второй обеспечивает отражение данных о местоположении объектов, их пространственной форме и ориентации (Лютый, 1981). Однако карты включают в свое содержание и естественноязыковые тексты — предложения, построенные из слов по правилам грамматики и

размещенные непосредственно на поле карты для дополнительного пояснения изображаемых явлений (Салищев, 1971). Поэтому правильнее говорить о языке карты как об искусственной знаковой системе передачи картографической информации, объединяющей: 1) систему условных знаков — собственно язык карты, 2) естественноязыковую терминологию пояснительных подписей и зарамочного оформления, 3) язык математической основы — картографическая проекция, масштаб и система координат, — определяющий пространственное расположение картографируемых объектов. Такое представление о языке карты соответствует реальной структуре картографической информации.

Требует разъяснения и интерпретация карты как текста. Действительно, с точки зрения семиотики функционирование языка карты в коммуникативных процессах позволяет рассматривать карты в качестве текстов — результата и средства коммуникации. Различие естественноязыковых и картографических текстов в том, что естественноязыковые тексты: 1) линейны, 2) воспринимаются как единая смысловая система, 3) требуют для понимания смысла текста последовательного чтения «от начала до конца», а карта: 1) двумерна, 2) читается от любой точки и в любом направлении, 3) порождает у читателя пространственный образ изображенного фрагмента действительности. Универсальные свойства языка карт лишь в последние четверть века стали предметом анализа: впервые этот анализ провел в 1967 г. французский ученый Жак Бертэн в классическом труде «Графическая семиология». В нашей стране проблемой семиотического и лингвистического анализа языка карт занимались А. Ф. Асланикашвили, А. С. Васмут, А. А. Лютый и другие ученые. Большой вклад в представление о картах как языковых коммуникативных моделях действительности внес замечательный советский ученый — картограф К. А. Салищев.

Для обеспечения оптимальной коммуникации содержание карты переводится на ЕЯ при помощи словаря, который в картографии носит название *легенды* карты. В этом словаре каждый картографический условный знак получает имя (название) на ЕЯ в виде термина или терминологического словосочетания (рис. 1).

Возникает неординарная семиотическая ситуация, когда один план содержания (объекты и явления действительности, подлежащие картографированию) взаимодействует с двумя планами выражения: системой условных знаков и знаками ЕЯ. При этом необходимо помнить, что ЕЯ является первичным по отношению к СУЗ как искусственной знаковой системе. Значения знаков генетически впервые были сформулированы в терминах ЕЯ и лишь на следующем этапе проектирования СУЗ соотнесены с конкретными знаковыми формами выражения. Генетическая связь между условным знаком и термином ЕЯ может утрачиваться, и искусственный язык может непосредственно

связываться с планом содержания (реальной действительностью), минуя естественноязыковый знак.

Условный знак	Название обозначаемых объектов
2.4	1) Аэродромы и гидроаэродромы 2) Участки автомобильных дорог оборудованные для взлета и посадки самолетов
1.2	1) Пристань с оборудованными причалами

Рис. 1.

Кроме условных знаков на картдокументе используются пояснительные подписи и другие термины ЕЯ (рис. 2).

Таким образом, язык карты — знаковая система, при помощи которой отображается предмет познания картографии — конкретное пространство предметов и явлений объективной действительности. Карту можно назвать текстом, представляющим собой сложную композицию условных знаков — графических фигур и знаков ЕЯ — *слов, словосочетаний, предложений*; при помощи этих знаковых систем передается информация об объектах местности: их расположении, количественных и качественных характеристиках.

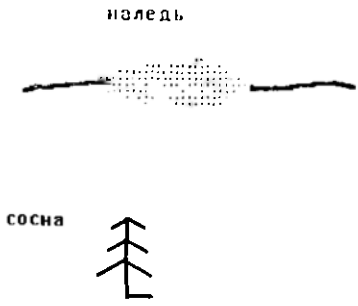


Рис. 2.

Основной единицей языка является слово. В языке карты

основной единицей является картографический условный знак (УЗ), репрезентирующий объект природного или социального

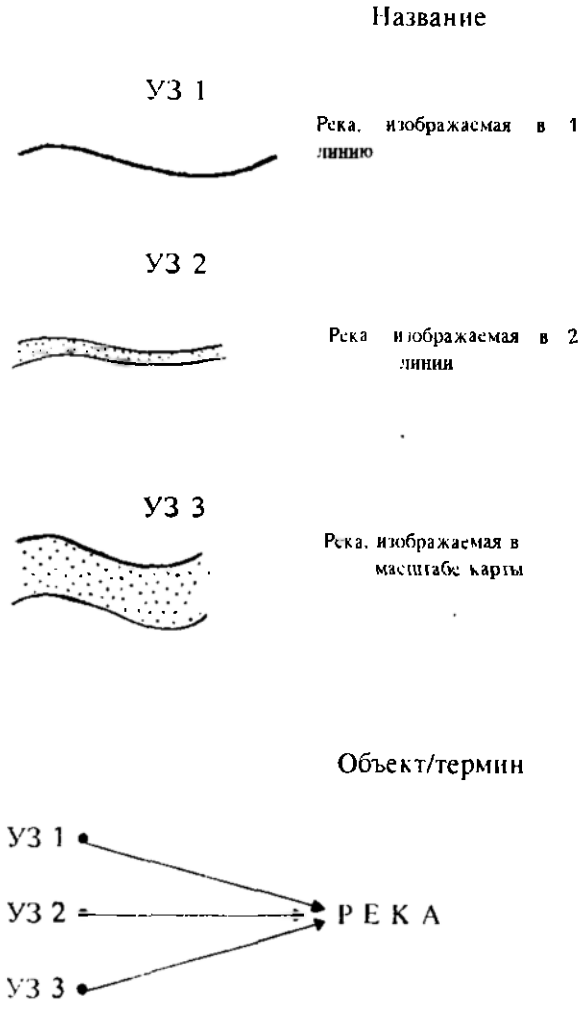


Рис. 3.

происхождения. Система УЗ реализует парадигматические и синтагматические отношения между языковыми единицами аналогично языку естественному.

Синонимия в системе УЗ хорошо иллюстрируется использо-

ванием разных знаков для передачи одного объекта действительности (рис. 3).

Полисемия в системе УЗ иллюстрируется фактами отображения одним УЗ нескольких объектов действительности (рис. 4).

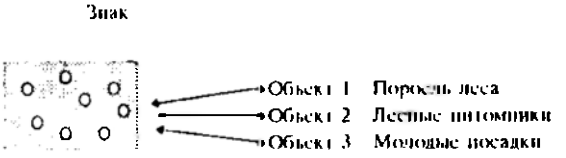


Рис. 4.

Гипонимические (родо-видовые) отношения широко распространены в системе УЗ (рис. 5).

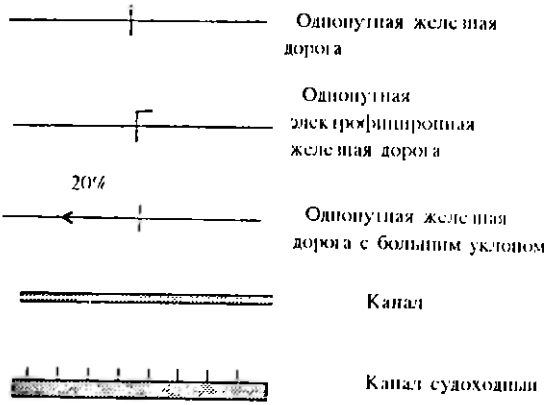


Рис. 5.

УЗ вступают в ситуативные (синтагматические) отношения, позволяющие передавать информацию о сложных объектах действительности. Например, сложные объекты «населенный пункт», «бассейн реки», «горный массив» и др. выражаются совокупностью УЗ «дом», «квартал», «обстановка» и т. д.; «река», «пристань», «порт», «приток», «устье» и т. д.; «подошва», «склон», «ледник» и т. д., вступающих в различные синтагматические отношения.

Рассмотренные выше отношения «работают» в основном на массивах лексических (словарных) единиц — наборах слов или УЗ. Анализируя картографические изображения как тексты, а язык карты как систему, выделяют синтактику, семантику, пра-

гматику и стилистику — фундаментальные свойства языковой системы.

Синтактика описывает внутреннюю структуру картографических изображений: форму, величину, цвет знаков, их комбинационные (синтагматические) возможности, процессы преобразования знаков при переходе от масштаба к масштабу (этот процесс называется генерализацией карты), правила построения знаков.

Семантика устанавливает отношения знаков к обозначаемым объектам. Эти отношения определяются, как говорилось выше, в языковой системе, а также в легенде карты, терминами ЕЯ. Картографический знак, как слово и словосочетание ЕЯ, называет предмет и выражает понятие, именно поэтому в языке карты возникают и существуют кратко рассмотренные выше отношения синонимии, полисемии, гиперонимии и другие семантические (смысловые) отношения.

Прагматика определяет правила применения знаков составителями карт, правила и условия восприятия знаков потребителями (читателями) карты. Главная задача прагматики — обеспечение оптимального понимания карты в зависимости от ее функционального назначения. Так, нет никакой необходимости, например, в штурманской аэронавигационной карте сохранять нагрузку, передающую мельчайшие особенности речной или почвенной системы.

Можно сказать, что синтактика устанавливает отношения знаков и элементов их внутренней структуры между собой, семантика — отношение знаков к обозначаемым объектам, прагматика — отношение знаковых текстов к потребителю карт, стилистика формирует подсистемы языка, уточняющие их функционирование в целях обеспечения оптимальной коммуникации — общения между человеком и знаковым текстом.

Стиль карты определяется как совокупность характерных черт карты, исходящих из целенаправленного выбора изобразительных средств. К субъективным стилеобразующим факторам относятся профессиональная эрудиция картографа, индивидуальные особенности составителей карты. Объективными стилеобразующими факторами можно считать систему изобразительных средств, технические средства, назначение карты. Различаются 4 группы стилей: 1) исторический (примитивный, античный, арабский, китайский и др.); 2) современный (прикладной, научный, популярный, художественный); 3) индивидуальный и издательский; 4) региональный и национальный. Стили карт изучаются в языковом и эстетическом аспектах. Стилистика карт основывается на традиционном признании карты произведением искусства.

В общем случае можно выделить следующие функции ЕЯ в картографических документах-текстах: 1) базовый язык для создания вторичной искусственной системы картографических

условных знаков; 2) самостоятельное коммуникативное средство для передачи картографического содержания.

Таким образом, карта как документ характеризуется двойственностью существующей лингвосемиотической ситуации: взаимодействием двух семиотических систем — терминологической системы ЕЯ и системы картографических УЗ. Вместе с тем все многообразие картографической информации (кроме пространственной локализации) может быть представлено в терминах ЕЯ, что оказывается совершенно необходимым при решении глобальных задач автоматизации создания и использования современных картографических систем на базе вычислительной техники.

В век электронно-вычислительной техники родилась новая область картографической науки и производства — автоматизированная картография, объектом которой является проектирование, разработка и эксплуатация автоматизированных картографических систем, или банков картографических данных.

Автоматизированная картографическая система (АКС) — банк данных и знаний, позволяющий решать следующие задачи:

— создание карт на основе геодезических данных и данных аэро-, фото- и космической съемки;

— создание карт производных масштабов на основе хранящихся в ЭВМ массивов карт базовых масштабов (генерализация);

— поисковые задачи, включая поиск информации о районах местности, прокладку авиационных и морских маршрутов, линий нефте- и газопроводов, электропередачи и др.

Выше отмечалось, что карты обладают сложной семантической структурой, которая складывается из:

— элементарных (простых) картографических объектов и характеристик, передаваемых УЗ языка карты (например, маяк, строение, шлюз, аэродром и т. д.);

— сложных объектов, образуемых сочетаниями УЗ, вступающих в смысловые связи, передаваемые системой грамматических отношений (река, бассейн реки, город и т. д.);

— предложений/высказываний, моделирующих сложные информационные ситуации (маршрут линии электропередач, навигационный курс самолета и т. д.).

Предложения — хотя и образованы из информационных единиц, называющих объекты различных типов, характеристики и отношения между ними, — непосредственно УЗ не выражены. Образование предложения происходит в результате мыслительной деятельности человека, читающего карту. В процессе образования предложений происходит решение логических задач, обеспечивающих получение нового знания.

Поэтому крупные промышленные банки картографических данных основываются на базах знаний, в которых хранятся све-

дения в цифровом виде, а также формализованные текстовые сообщения на ЕЯ.

Под знанием понимаются сведения, отображающие закономерности области знания и позволяющие выводить новые факты и прогнозировать развитие отрасли знания. Для построения базы знаний должна быть создана информационно-лингвистическая модель отрасли знания (предметной области), т. е. построена совокупность возможных и допустимых состояний предметной области, выраженных через термины ЕЯ. Такие модели называются в информатике концептуальными, или инфологическими, моделями и содержат полное информационное описание объектов и отношений между ними. Модель обычно задается в виде лингвистического представления — словаря или классификатора. Модель передает сложные семантические связи между объектами, обеспечивающие представление сложных картографических объектов и связей между ними, а также сложных информационных ситуаций, что позволяет решать задачи логического преобразования информации, получать ответы, оценивать ситуации.

АКС состоит из подсистем:

— ввода информации с помощью устройств снятия метрической информации путем ручного отслеживания или автоматического сканирования;

— семантической подготовки и обработки информации в виде классификатора понятий/терминов предметной области;

— хранения информации в ЭВМ;

— программного комплекса преобразования информации;

— вывода информации на графопостроитель или печатающее устройство.

Для построения АКС, реализующих базы знаний, проектируются специальные лингвистические процессоры, которые могут структурироваться в подсистеме семантической подготовки и обработки информации (подсистеме информационно-терминологического обеспечения).

Информационно-терминологическое обеспечение — это средство идентификации, описания, структурирования и кодирования единиц информации, представленных в виде понятийно-терминологической системы предметной области и системы УЗ и ориентированных на автоматизацию процедур ввода, хранения и обработки картографической информации.

Рассмотрим функциональные связи и определим основные блоки в подсистеме информационно-терминологического обеспечения (рис. 6). Его базовая структура — это структурная модель предметной области, представляющая объекты и характеристики картографии в терминах ЕЯ, отношения между ними, а также систему УЗ.

Классификатор картографической информации — множество иерархически упорядоченных картографических объектов и их

группировок, картографических характеристик с выбранной системой кодирования, а также множество иерархически упорядо-

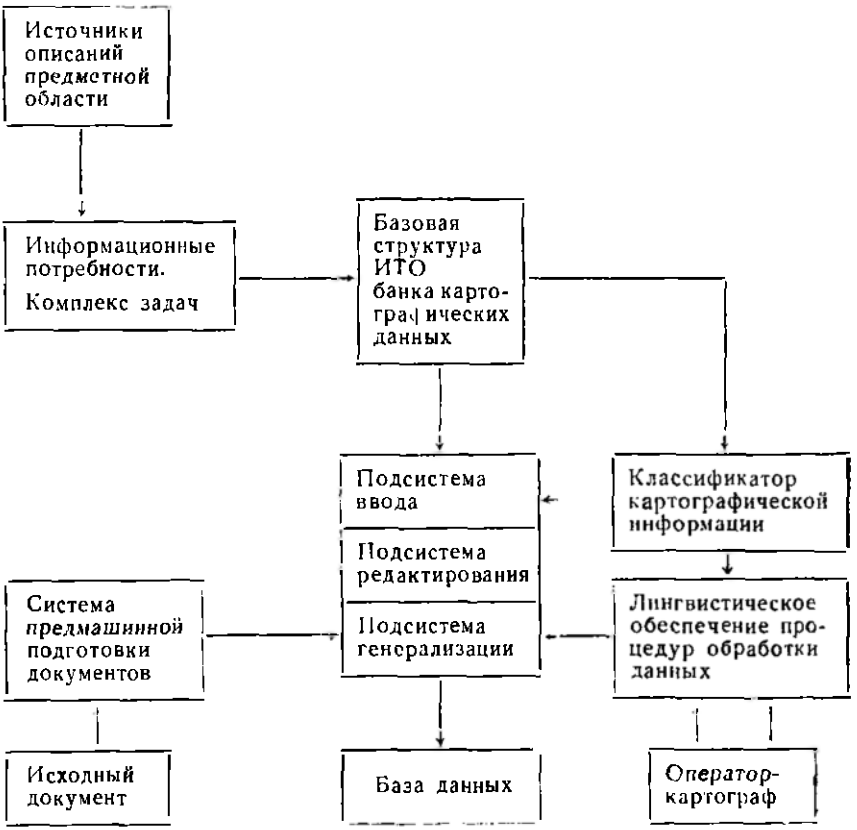


Рис. 6.

ченных графических элементов системы УЗ. Приведем фрагмент классификатора терминов понятий:

КЛАССИФИКАЦИЯ ОБЪЕКТОВ 2 РАЗДЕЛА

СПРАВКА	КОД	ПОНЯТИЕ
B15C34444351	2	ЭС «ГИДРОГРАФИЯ И ГИДРОТЕХНИЧЕСКИЕ СООРУЖЕНИЯ»
B3 I	2.1.	ГИДРОГРАФИЯ
B16C54444351	2.1.1.	РЕКИ И РУЧЬИ
B16C54444351	2.1.1.1	РЕКИ И РУЧЬИ ПОСТОЯННЫЕ
B13311433351	2.1.1.1.1	РЕКА И РУЧЕЙ ПОСТОЯННЫЕ.ВЫР.В М.К.
A12232433111	2.1.1.1.2	РЕКА И РУЧЕЙ ПОСТОЯННЫЕ В 2 ЛИНИИ
B12232433111	2.1.1.1.3	РЕКА И РУЧЕЙ ПОСТОЯННЫЕ В 1 ЛИНИЮ

A12232433121	2.1.1.1.4	КАНАЛИЗОВАННЫЙ УЧАСТОК РЕКИ
B16C54444351	2.1.1.2	РЕКИ И РУЧЬИ ПЕРЕСЫХАЮЩИЕ
A12232433111	2.1.1.2.1	РЕКА И РУЧЕЙ ПЕРЕСЫХАЮЩИЕ В 2 ЛИНИИ
A12232433111	2.1.1.2.2	РЕКА И РУЧЕЙ ПЕРЕСЫХАЮЩИЕ В 1 ЛИНИЮ
B16C54444351	2.1.1.3	ПОДЗЕМНЫЕ И ПРОПАДАЮЩИЕ УЧАСТКИ РЕК
812212433111	2.1.1.3.1	ПОДЗЕМНЫЙ И ПРОПАДАЮЩИЙ УЧАСТОК РЕКИ В 2 ЛИНИИ
B12232433111	2.1.1.3.2	ПОДЗЕМНЫЙ И ПРОПАДАЮЩИЙ УЧАСТОК РЕКИ В 1 ЛИНИЮ
B12321444351	2.1.2	ОЗЕРО
B13211441151	2.1.3	МОРЕ
B3 1	2.1.4	ОБЪЕКТЫ ПРИБРЕЖНОЙ ПОЛОСЫ
B13131444351	2.1.4.1	БЕРЕГОВАЯ ОТМЕЛЬ И МЕЛЬ
9 13111444351	2.1.4.2	БЕРЕГ ОПАСНЫЙ (ХАРАКТЕР ОПАСНОСТИ НЕИЗВЕСТЕН)
B13154444351	2.1.4.3	БЕРЕГА ОСЫХАЮЩИЕ (ПРИЛИВНО-ОТЛИВНЫЕ ПОЛОСЫ)
B13131444351	2.1.4.3.1	БЕРЕГ ОСЫХАЮЩИЙ ПЕСЧАНЫЙ
813131444351	2.1.4.3.2	БЕРЕГ ОСЫХАЮЩИЙ ИЛИСТЫЙ
813131444351	2.1.4.3.3	БЕРЕГ ОСЫХАЮЩИЙ ПЕСЧАНО-КАМЕНИСТЫЙ И ГАЛЕЧНО-ГРАВИЙНЫЙ
813131444351	2.1.4.3.4	БЕРЕГ ОСЫХАЮЩИЙ СКАЛИСТЫЙ

Лингвистическое обеспечение процедур обработки данных — это язык диалога «человек — ЭВМ». ЭВМ предъявляет оператору меню, построенное на основе базовой структуры и классификатора, из которого выбирает сообщения, описывающие тот или иной объект карты:

DK3 : [1.1] ТОРМАР
ВВОД ОБЪЕКТА
ОПРЕДЕЛИТЕ РАЗДЕЛ
1. ЭС «ОПОРНЫЕ ПУНКТЫ»
2. ЭС «ГИДРОГРАФИЯ И ГИДРОТЕХНИЧЕСКИЕ СООРУЖЕНИЯ»
3. ЭС «НАСЕЛЕННЫЕ ПУНКТЫ (НП)»
4. ЭС «ПРОМЫШЛЕННЫЕ, СЕЛЬСКОХОЗЯЙСТВЕННЫЕ И СОЦ.-КУЛЬТ. ОБЪЕКТЫ»
5. ЭС «ДОРОЖНАЯ СЕТЬ»
6. ЭС «РЕЛЬЕФ»
7. ЭС «РАСТИТЕЛЬНОСТЬ И ГРУНТЫ»
8. ЭС «ГРАНИЦЫ»
9. ОПИСАНИЕ ЛИСТА
0. СИСТЕМНЫЙ РАЗДЕЛ

2

ОПРЕДЕЛИТЕ
ЭС «ГИДРОГРАФИЯ И ГИДРОТЕХНИЧЕСКИЕ СООРУЖЕНИЯ»
1. ГИДРОГРАФИЧЕСКИЕ ОБЪЕКТЫ
2. ГИДРОТЕХНИЧЕСКИЕ СООРУЖЕНИЯ
1

ОПРЕДЕЛИТЕ
ГИДРОГРАФИЧЕСКИЕ ОБЪЕКТЫ
1. ВОДНЫЕ ОБЪЕКТЫ
2. БЕРЕГ
3. ЕСТЕСТВЕННО-ГИДРОГРАФИЧЕСКИЙ РЕЛЬЕФ
4. ВОДНЫЕ ПРЕГРАДЫ

Система предмашинной подготовки документов — система, в которой картографическая информация переводится уже в форму данных для последующего ввода. При этом: 1) из документа выделяется исходный картографический объект; 2) объектам присваиваются порядковые номера для дальнейшей идентификации объектов.

Современные банки картографических данных лежат в основе систем электронных карт, которые широко используются в авиационной и морской навигации.

Литература

Асланикашвили А. Ф. Метакартография. Тбилиси, 1974.

Васмут А. С. Моделирование в картографии с применением ЭВМ. М., 1983.

Куприн А. М. Слово о карте. М., 1987.

Лютый А. А. Язык карты. М., 1981.

Салищев К. А. Картоведение. М., 1976.

СТИЛЕМЕТРИЯ

ФИЛОЛОГИЧЕСКИЙ СТАТУС СТИЛЕМЕТРИИ

Стилеметрия — прикладная филологическая дисциплина, занимающаяся измерением стилевых характеристик с целью систематизации и упорядочивания (типологии, атрибуции, датировки, диагностики, реконструкции и т. п.) текстов и их частей.

Термин «стилеметрия» ввел в научный обиход немецкий филолог В. Диттенбергер, решавший в конце прошлого столетия проблему авторства диалогов Платона с помощью частот слов, реализация которых не зависит от тематики текста (к таковым Диттенбергер относил в основном служебные слова) (Dittenberger, 1880). В России первым «стилеметром» был народоволец Н. А. Морозов, известный писатель и разносторонний ученый. Он, как и Диттенбергер, использовал частотные распределения (спектры) служебных слов для отличения плагиата от подлинных произведений (Морозов, 1915).

К настоящему времени накоплен большой опыт в применении статистического метода для решения проблемы спорного авторства (атрибуции) текстов; в сущности, этот метод стал хрестоматийным, причем не только в филологии (текстологии, источниковедении, стилистике), но и в криминалистике, медицинской диагностике, социальной психологии. Примечательно, что в самой статистике атрибуция текста стала выступать в кругу типовых задач, иллюстрирующих широту исследовательских возможностей статистики как метода научного познания (Кимбл, 1982; Кокс, Снелл, 1984).

В последние годы круг решаемых стилеметрией задач и репертуар применяемых ею методов существенно расширились. Практической повседневностью стала количественная таксономия текстов (Krous, Vašak, 1967; Doyle, 1965; Шайкевич, 1980; Тулдава, 1981; Бородкин, 1981), стилистическое приложение нашли дешифровочные модели (Тимофеева, 1983), относительно само-

стоятельное направление образовала квантитативная типология текста (Алексеев, 1983), начала формироваться стилистическая диагностика (Статистичні параметри стилів, 1967; Севбо, 1981), большое развитие получили методы реконструкции древних текстов (Бородкин, 1986), продолжали развиваться традиционные стилестатистические методы.

ИЗМЕРЕНИЕ СТИЛЯ

Под измерением чаще всего понимается процесс определения отношения измеряемой величины к другой однородной величине, которая принята за единицу. Главным в трактовке понятия измерения является то, что оно соотносится с объектом исследования через свойства и характеристики, которым может быть приписана определенная величина. Однако вправе ли мы говорить о величинах (и соответственно об измерении) там, где *царит исключительная, рафинированная качественность*, на первый взгляд, не поддающаяся любой разумной квантификации? Можно ли, например, измерить сложность текста, его художественность, целостность, оригинальность, образность, метафоричность, можно ли измерить степень сходства между текстами, оценивать диагностическую существенность стилистического признака и т. п.? Утвердительный ответ на некоторые из этих вопросов, конечно, возможен, но лишь в том случае, когда удается выявить в исследуемом явлении элементарные, доступные измерению признаки, которые в совокупности, интегрально могут дать объективную картину в виде меры и числа. Стиль, строго говоря, нельзя непосредственно измерить, так же как нельзя измерить счастье и творческие способности, его можно только описать. Об измерении можно говорить только тогда, когда исследователю удается разумно атомизировать это сложное явление и каждому элементарному значению признака приписать соответствующую величину.

Кроме наук физического цикла измерениями в настоящее время пользуются почти все естественные и общественные дисциплины, причем в некоторых из них сложились специализированные ответвления, занимающиеся измерением: биометрия, антропометрия, социометрия, наукометрия и др. К этой группе научных дисциплин можно отнести и стилеметрию.

Наиболее тесные родственные узы объединяют ее с теми метрическими дисциплинами, для которых текст в различных аспектах своих внутренних и внешних связей является исходной данностью и единым объектом исследования. К их числу относится группа научных дисциплин гуманитарного профиля, занимающихся разысканием измеримых параметров культуры (научной, художественной, массовой, технологической) в потоках информации, циркулирующих в виде текстов, которые образуют «память мира» или «память культуры». Такими тексто-

измеряющими дисциплинами являются: наукометрия, социометрия, искусствометрия, клиометрика. Без большой натяжки можно сказать, что социолог и науковед часто выполняют стилеметрическую, а лингвист — наукометрическую или социометрическую работу.

ОБЪЕКТ И ПРЕДМЕТ СТИЛЕМЕТРИИ

Объектом стилеметрии является текст, но это не текст вообще, это всегда конкретный текст, т. е. текст, созданный конкретным автором, в конкретное время, в конкретной ситуации. При этом текст мыслится как единичное собирательное понятие, в котором отражены признаки группы предметов, образующих единое целое.

Понятие «текст», впрочем, как и любое другое понятие, может быть разделительным и собирательным, в зависимости от того, как оно рассматривается. Так, «текст» есть собирательное понятие относительно входящих в него словоупотреблений, но он — разделительное понятие относительно различных видов текста (научных, художественных, деловых и прочих текстов).

С точки зрения теории множеств объект стилеметрии — собирательное множество, а с точки зрения теории систем текст может быть отнесен к классу внутренних систем, являющихся целостными образованиями, к которым можно применять процедуры членения, представляя эти системы в виде некоторой структуры составляющих их частей. Для внутренних систем характерно то, что здесь целостность доступна прямому наблюдению, имеет пространственно-временную ориентацию (Шрейдер, Шаров, 1982; Чебанов, 1980).

С точки зрения теории статистики текст может рассматриваться как реальная совокупность. Эта категория была впервые введена А. А. Чупровым применительно к различного рода реально существующим, а не создаваемым мыслью исследователя сообществам и коллективам, представляющим собой целостное образование сосуществующих и взаимодействующих единиц, не обязательно однородных. Основным признаком таких реальных совокупностей А. А. Чупров считает их устойчивость во времени: способность сохранять в течение более или менее длительного времени свой состав и характерные черты (Чупров, 1910). С этой точки зрения текст сверхустойчив: ни одно слово, ни одна фраза из текста удалена быть не может.

Собирательным понятиям в логике противопоставляются разделительные понятия, в математике соответственно разделительные множества, а в теории систем — внешние системы. Элементы, объединяемые при этом в множестве, соотносимом с разделительным понятием, в отличие от собирательного множества рассматриваются как неразличимые. Логические, мате-

матические и филологические аспекты собирательных и разделительных категорий представлены в таблице.

Методологический статус и содержание собирательных и разделительных категорий

Методологический статус	Категория	
	собирательная	разделительная
Понятие	Собирательное	Разделительное
Объем понятия	*	Класс референтов
Соотношение объема и содержания понятия	*	Обратное
Процедура выделения частей	Членение	Разбиение
Задача индукции по неполному основанию	Реконструкция	Экстраполяция
Характер составляющих	Разнородные	Однородные
Представление понятия в виде системы	*Внутренняя система*	*Внешняя система*
Характер системы	Пространственно-временная организация	Абстрактно-логическая конструкция
Совокупность составляющих	Собирательное множество	Разделительное множество
Модель	Мереология Лесневского	Множества Кантора
Теория статистики	Естественная совокупность	*Искусственная* совокупность
Филологические примеры	Текст Русская литература XIX в.	Словник Список произведений

* *В классической логике не определен.*

Теперь перейдем к выяснению того, какие структуры являются предметом стилеметрии.

В современной методологии науки принято описывать структуру объекта как архетип (Мейен, Шрейдер, 1976). В последнем можно вычленить мероны (*части, элементы, свойства, отношения*), осознаваемые как признаки, среди которых можно выделить существенные. Под существенными признаками в логике понимаются такие признаки, после изъятия которых предмет перестает быть самим собой. Скажем, если мы лишим равносторонний треугольник равенства сторон, то лишимся и его самого, так как получим разносторонний треугольник. Существенным признакам будут сопутствовать собственные и несобственные неотделимые признаки. Все они отражают обязательные, характерные мероны архетипа, логически (включая вероятностную логику) выводимые из существенных признаков. Кроме того, объект характеризуется и несобственными отделимыми признаками, которые практически независимы от существенных меронов и отвечают меронам, факультативно присутствующим в архетипе. Совокупность последних можно назвать периферией характеристики объекта. Тогда под элемента-

ми стиля можно понимать особенности периферии характеристики объекта (например, для стула обязательным мероном будет мерон «ножки», который может быть в состоянии «присутствует — отсутствует», а факультативными состояниями будут модальности «количество», «цвет», «какого дерева» и т. п.). Другими словами, стиль может быть описан через факультативные, поверхностные признаки объекта, которые лишь неявным образом затрагивают сущностные, глубинные характеристики.

КЛАССИФИКАЦИОННЫЕ ЗАДАЧИ СТИЛЕМЕТРИИ

Стилеметрия имеет дело с количественным классифицированием, а эта область классификационных занятий тесно соприкасается с несколькими научными направлениями: теорией группировок, теорией оценивания, распознаванием образов, теорией корреляции, количественной таксономией, методами психологического тестирования и др. В настоящее время границы между этими направлениями стираются, и можно говорить о комплексе подходов и методов, занимающихся теми или иными видами количественной систематизации объектов произвольной природы.

Предлагаемый анализ основан на различении собирательных и разделительных категорий и связанных с ним процедур членения и разбиения.

Всю совокупность видов стилеметрической работы (типология текста, атрибуция текста, стилистическая диагностика и др.) можно обозначить как упорядочивающе-систематизирующую деятельность. Она распадается на две большие области: упорядочивание и систематизацию (Мартыненко, Чебанов, 1988).

Формы упорядочивания. Когда исследователь имеет дело с упорядочиванием, его интересует только содержание соответствующих понятий, соотношение этих понятий, а также вопрос о том, соотносима ли характеристика данного референта (например, текста или его части) с содержанием соответствующего понятия. При этом мы можем отвлечься от того, что такой референт не единствен, а существует большее или меньшее количество аналогичных референтов. Такая деятельность, в свою очередь, многообразна, и можно говорить о разных ее задачах.

Начнем с задачи диагностики. Занимаясь ею, мы должны перечислить существенные свойства того или иного объекта, обладание которыми позволяет нам идентифицировать объект с рассматриваемым понятием. Для этого необходимо описать исследуемое явление, выявить зависимости, характеризующие его структуру. Часто сделать это строгими методами классической логики не удается ввиду того, что рассматриваемое явление оказывается достаточно многообразным

и в нем практически невозможно выявить какие-либо устойчивые признаки, среди которых будут находиться и претендующие на то, чтобы быть существенными. В этом случае мы можем охарактеризовать такое понятие некоторым диагностическим синдромом — совокупностью симптоматических признаков, которые чаще всего присущи рассматриваемому явлению. В качестве «измерительного инструмента» здесь используются различные методы, направленные на отбор прагматически существенных (полезных) диагностирующих признаков из множества априорно заданных симптоматических признаков (факторный анализ, метод максимального корреляционного пути, метод корреляционных плеяд и др.).

Примером решения диагностической задачи являются работы Г. Ф. Мальцевой и М. А. Марусенко, в которых осуществляется оптимальная свертка исходного признакового пространства художественного текста с целью выделения наиболее информативных диагностирующих признаков (Мальцева, 1969; Марусенко, 1987).

Важно обратить внимание на то, что при решении диагностических задач работа производится с признаками и их связями, без учета того, какие именно объекты и их совокупности являются носителями конкретных сочетаний признаков. Таким образом, упорядочивание здесь осуществляется в контексте собирательных понятий.

На основе решения диагностической задачи мы можем идентифицировать тот или иной конкретный объект, отождествить его с диагнозированным понятием. Пусть, например, мы установили систему признаков, характеризующих художественную прозу, и нам предъявлен текст, ранее нам неизвестный. Мы начинаем выяснять, обладает ли данный текст тем же набором признаков. Выяснив это, мы можем квалифицировать данный текст как относящийся или не относящийся к художественной прозе.

Возможна и другая ситуация. Мы сталкиваемся с некоторым явлением, но не можем охарактеризовать его через набор явных признаков, чтобы соотнести его с каким-либо понятием. И тем не менее мы можем исследовать данное явление, но для этого ему нужно дать какое-то имя. Тогда отождествить это явление с тем же явлением, встретившимся повторно, мы можем, употребив то же самое имя, т. е. квалификация в данном случае осуществляется через и м е н о в а н и е.

Такой способ квалификации особенно важен при изучении малоизвестных литературоведам или лингвистам явлений, когда исследователь выступает в качестве эксперта, скорее чувствующего интересующее его явление, чем способного его явно описать и объяснить. Вместе с тем результат экспертной оценки может быть предметом статистической обработки. К числу наименований-оценок можно отнести, например, качественные

характеристики стиля (выразительность, эмоциональность, напряженность и т. п.), с помощью которых экспертами-информантами оцениваются достоинства литературных произведений (Кэрролл, 1972).

Результаты диагностики и именования часто используются не изолированно, а соотносятся друг с другом. Это находит свое отражение в построении шкал. Важно отметить, что, осуществляя шкалирование, мы имеем дело не с конкретными объектами, совокупности которых являются разделительными множествами, а с содержаниями (интенсионалами) разных понятий. Начиная соотносить эти содержания друг с другом, мы обращаемся с ними как с индивидами, но способ существования таких индивидов — индивидов-архетипов — совсем иной, чем у индивидов — эмпирических референтов. В частности, в то время как эмпирических референтов может быть много, архетип существует в единственном «экземпляре». Таким образом, упорядочивание и здесь осуществляется в контексте собирательных понятий.

Существует, как известно, три типа шкал: наименований, порядковая и количественная. Включая название в шкалу, давая объекту наименование, мы находим ему место в номинальной шкале. Ранжируя диагностические характеристики, мы получаем порядковую шкалу, а присваивая им числа, измеряя их, мы находим место объекта на количественной шкале. Важно отметить, что положение на шкале может быть зарегистрировано для каждого отдельно взятого объекта. Например, взяв предложение из рассказов А. П. Чехова, можно найти его место на количественной шкале размера предложения.

Итак, различные виды упорядочивания — диагностика, именование, шкалирование — позволяют по-разному квалифицировать (идентифицировать, назвать, наименовать, ранжировать, измерить) отдельно взятый объект, получив в результате то, что понадобится для систематизации их совокупностей.

Формы систематизации. Рассмотрим типы деятельности, в которых работа с собирательными понятиями переплетается тем или иным способом с работой с разделительными понятиями. Здесь имеется в виду следующая ситуация: даже тогда, когда мы ориентируемся на результат, формируемый с помощью собирательных понятий, нужно все время иметь в виду и работу, сопряженную с разделительными понятиями.

Фрагментация представляет собой процедуру разделения целого на фрагменты. При этом целое может быть представлено как собирательным, так и разделительным множеством, и соответственно разделение здесь выступает как членение и разбиение. Оба типа разделения могут производиться на основании параметров, которые принимают разные значения на определенных фрагментах целого или на определенной таксономической совокупности, мыслимой как разделительное

множество. В этом случае мы имеем дело с параметрической фрагментацией. Например, конкретное множество текстов может быть разбито на отдельные фрагменты на основании параметра «объем текста». Этот объем может быть измерен количеством словоупотреблений, предложений, абзацев или каким-либо другим способом.

Членение и разбиение при фрагментации могут образовывать композиции.

Разделения, композиции которых коммутативны, дают комбинативные (фасетные) структуры представления строения или таксономических отношений. Когда мы рассматриваем текст как собирательное понятие, то в нем можно выделить куски, соответствующие авторской речи и речи персонажей и далее в каждом из полученных фрагментов выделить служебные и знаменательные слова. Можно произвести то же самое расчленение в обратном порядке и получить аналогичное итоговое расчленение. Точно так же все произведения русской художественной литературы как разделительное понятие можно разбить на произведения, относящиеся к разным эпохам и разным жанрам в любой последовательности. В этом случае можно говорить о фасетной фрагментации.

Разделения, композиции которых не коммутативны, свойственны иерархическим структурам, в которых строение целого или таксономический универсум последовательно разделяются на фрагменты, делимые, в свою очередь, на более мелкие фрагменты. Например, совокупность текстов на русском языке может быть разбита на тексты, относящиеся к различным функциональным стилям. После этого тексты, относящиеся, скажем, к художественному стилю, могут быть разбиты на тексты, принадлежащие к различным жанрам (поэзия, проза, драма и др.) и т. п. Такой тип фрагментации можно назвать иерархическим.

Следующий тип систематизации — морфологизация. В этом случае мы вычленяем и описываем обобщенное устройство (структуры) объектов некоторой группы (не интересуясь, вообще говоря, тем, как организованы сами эти группы). Описывать организацию можно по-разному. Так, если мы описываем организацию в открытых параметрах и реконструируем латентные характеристики как некоторые инварианты, то морфологизация выступает как структурализация; если же морфологизация описывается более богато, но менее конструктивно, то выделяются обобщенные типы организации — архетипы, исследованием которых занимается мерономия (Гете, 1975). Обращаясь к форме как к единству плана выражения и плана содержания, т. е. интерпретируя ее как нечто содержательное, мы имеем дело с собственно морфологией, которая, правда, в большой мере лишается конструктивности. Занимаясь морфологией в том или ином смысле, мы выявляем сово-

купности частей и их взаимосвязей, присущих объектам данной группы. Применительно к языкознанию структурализация может быть соотнесена с математическими моделями языка (например, грамматикой непосредственно составляющих или грамматикой зависимостей), мерономия — с традиционным синтаксисом, а морфология — с грамматическими учениями античных философов.

По характеру работы морфология сходна с упорядочиванием. Однако между ними есть и принципиальные различия. Так, при упорядочивании не принимается во внимание множество существующих референтов. Для морфологии это существенно. Кроме того, упорядочивание ориентировано на выделение жестких границ между значениями признаков, в то время как для морфологии актуален факт полиморфизма референтов. Это различие существенно, в частности, для выделения текстов как индивидов. В большинстве случаев, особенно при работе с современными текстами, эта задача для текстолога-эмпирика является тривиальной, однако при разыскании формальных критериев, позволяющих квалифицировать конкретную последовательность предложений как текст, возникают серьезные затруднения. Так, в лингвистике текста изучаются маркеры начала и конца текста, статистические методы используются для установления целостности текста, его отличия от фрагмента текста или конгломерата нескольких текстов. Например, Ю. К. Орлов, изучая роман Л. Н. Толстого «Война и мир» на предмет соответствия распределения словоформ закону Ципфа (см. раздел «Методы статистического моделирования в языкознании»), приходит к выводу, что текст этого романа следует рассматривать как совокупность нескольких текстов, каждый из которых подчиняется закону Ципфа, в то время как текст всего романа этому закону не удовлетворяет (Орлов, 1970). Выделение индивидов — важнейший вид работы по морфологизации и одновременно один из наиболее сложных компонентов предклассификационной деятельности. Принципиально, что при этом выделяется не один индивид, а сразу класс индивидов.

Основным видом задач индукции по неполному основанию является задача реконструкции — выявление элементов и связей между ними, которые не даны исследователю в наблюдении. Филологическим вариантом решения такой задачи являются дешифровочные алгоритмы, извлекающие грамматику из текста (Сухотин, 1963). Результаты работы таких алгоритмов могут быть использованы для стилистической диагностики текстов (каждый текст имеет «свою» грамматику) (Тимофеева, 1983).

Особым случаем морфологизации является районирование, при котором осуществляется расчленение (мерономизация) пространства. Специфика районирования как фор-

мы мерономизации заключается в том, что разные части формы мы привязываем к разным частям пространства, внеположенного, вообще говоря, этой форме. Классической лингвистической задачей на районирование является выделение диалектных ареалов в лингвогеографии, в том числе с использованием статистических методов. Сходный характер имеет также работа по выделению текстовых ареалов в ареальной типологии текстов (Герд, 1982).

Как районирование в пространстве признаков можно рассматривать некоторые подходы к классифицированию. В этом случае классифицируемый объект мы представляем как набор значений признаков, задающих точку в многомерном фазовом пространстве. Районы в пространстве признаков задают соответствующие таксоны. Выделением таких районов занимается дискриминантный и кластерный анализ. Дистрибутивно-статистический метод решает задачу такого же типа: с его помощью на основе информации о размещении единиц в тексте осуществляется количественная таксономизация текстов или единиц текста (Шайкевич, 1980).

Как специфический тип районирования можно трактовать районирование во времени — периодизацию и связанную с ней датировку. Очевидно, что периодизация также является формой работы с собирательными понятиями. Это своего рода временна́я морфологизация. Периодизация — самостоятельная и логически равномощная морфологии сфера деятельности, поскольку первая занимается квазивременными, а вторая — квазипространственными аспектами строения объектов.

В связи с проблемами временно́го (исторического) расчленения ведущими задачами индукции по неполному основанию являются исторические реконструкции, причем ввиду направленности времени можно говорить о двух типах реконструкций: ретрогнозах и прогнозах. При решении этих задач существенно выявить, о каком времени идет речь. Так, поток речи существует в одном времени, и в этом случае перед нами может стоять задача восстановления предшествующих или предсказания последующих отрезков текста. Такая задача часто решается в текстологии. Другой смысл имеет время, в котором происходит эволюция индивидуального стиля. В этом случае ретрогноз и прогноз будут соотноситься с уже написанными произведениями и теми, которые еще предстоит написать данному автору.

Говоря о периодизации, важно отличать ее от временны́х классификаций. Дело в том, что произведения одного и того же автора, написанные в один период, могут иметь сходство друг с другом и принадлежать ввиду этого к некоторому таксону. Однако совокупность произведений — разделительное по-

нятие, в то время как соответствующий ему период творчества — понятие собирательное.

Следующий тип систематизации — группирование, основная задача которого — разнесение объектов по группам. Само собой разумеется, что интерес представляют только такие группы, которые включают в себя объекты, сходные в каком-то нетривиальном отношении. Это равносильно тому, что с данной группой соотносится достаточно богатый архетип. Таким образом, задача группирования неотделима от задачи морфологизации, и по характеру последней, с учетом особенностей самих групп, можно выделить разные типы группирования.

Прежде всего можно говорить о параметризации, в которой архетип задается фиксацией одного или нескольких параметров, и все объекты, удовлетворяющие фиксированному набору параметров, относятся к данному классу. Очевидно, что архетип в этом случае достаточно беден. Однако, работая с измеряемыми параметрами, можно в принципе разбить генеральную совокупность на систему непересекающихся классов. Примером такой работы будет разнесение словоупотреблений какого-то текста на классы по числу графем.

Собственно классификация — это такой способ систематизации (группирования), когда класс объектов выделяется на основе некоторых предварительно выделенных существенных признаков. Богатство классификации зависит от того, насколько удачно выбраны существенные признаки. Чтобы охватить все множество объектов, при классифицировании целесообразно следовать закону единства основания деления понятий. Такой принцип строго выдерживается в исключительных случаях, например, в документных классификациях иерархического типа. Для лингвистики и литературоведения более характерны типологии, являющиеся еще одним способом группирования. В основу выделения типов кладется выявление важных для конструирования из них идеальных образов — типов как способов представления содержания конструируемого понятия. Сконструированный тип может реализоваться в относительно чистом виде, и тогда мы говорим о типических представителях, или один объект может сочетать в себе свойства нескольких типов. Если свойства каждого типа выражены при этом неявно, то мы имеем дело с переходными формами; в том же случае, когда в одном объекте выражены свойства нескольких типов, возникает смешанный (сложный) тип. Очевидно, что чистые и сложные типы являются вариантами архетипа в собственно типологии. Спецификой типа в интенсиональном (понятийном) аспекте является то, что это смысловой, семантический инвариант, а не количественный (как в параметризации) или структурный (как в классификации). Кроме того, для типологии характерно то, что в интенсиональном ас-

пекте мы рассматриваем не только тип, но и возможные его варианты, и в идеале, описывая тип, мы должны описать и совокупность его вариантов, а также пределы их варьирования. Однако это удается далеко не всегда. Особенностью экстенсионального аспекта типологии, как следует из сказанного, является то, что группа, ассоциируемая с данным типом, принципиально неоднородна — в ней есть как характерные, типические представители, так и периферийные, чем-то отличающиеся от типических. Кроме того, в типологизируемом универсуме могут оказаться объекты, которые вообще нельзя связать с каким-либо типом.

Итак, занимаясь типологией, в принципе мы не производим разбиения универсума объектов на классы, а получаем, скорее, частично пересекающиеся созвездия наряду с изолированными элементами. Примером типологизации может служить разделение совокупности текстов (или их авторов) на группы по критерию стилистической близости: бо́льшая часть текстов образует несколько многоэлементных типических групп, часть из которых пересекается, а некоторые тексты остаются изолированными (те, которые отличаются оригинальностью стиля).

Для всех форм группирования основной задачей индукции по неполному основанию является задача экстраполяции (таксономической и типологической). Суть ее заключается в том, что результаты, полученные на одной части таксономического универсума, переносятся на другие его части или на универсум в целом.

Взаимосвязи форм упорядочивания и систематизации. Описанные выше формы упорядочивания и систематизации на практике никогда не встречаются в изолированном виде: существование каждой из них невозможно без существования других.

Рассмотрим теперь несколько примеров взаимодействия форм упорядочивания и систематизации в стилеметрических исследованиях.

Пусть мы имеем какую-нибудь удачную и содержательную типологию функциональных стилей. Изучив некое произведение, мы должны отнести стиль этого произведения к какому-то типу упомянутой типологии. Для этого мы должны располагать некоторой процедурой отождествления стиля данного произведения с одним из выделенных ранее типов стилей, т. е. осуществить идентификацию. Но идентификация, как говорилось выше, является видом квалифицирования как формы упорядочивания, т. е. идентификация связывает систематизацию и упорядочивание. В филологии такая задача фигурирует как атрибуция. Если мы имеем дело с группировкой стилей на основе некоторых измеримых параметров (т. е. занимаемся параметризацией), то отнесение к

группе будет осуществляться с помощью измерения, а сами универсумы интенсиональных групп, выделяемых при параметризации, слагают шкалу. Такой вид работы относится к компетенции стилеметрии.

Иллюстрацией другого типа работы будет квантитативная типология текста, по П. М. Алексееву (1983). В этом случае практически исследуется морфология текста определенного вида литературы с соответствующим набором жанров. При этом морфология выступает в виде структурализации, причем особенности структуры (преимущественно лексико-семантической) описываются параметрами, что позволяет использовать методы квантитативной лингвистики. Такой способ работы полностью относится к оперированию собирательными понятиями. Однако для того чтобы строить структуру абстрактного текста, приходится учитывать некоторую типологию текстов, используя в качестве генеральной совокупности группу текстов одного типа (например, научно-технические тексты, публицистические, эпистолярные и др.). Таким образом, осуществление морфологизации без хотя бы тривиального группирования невозможно.

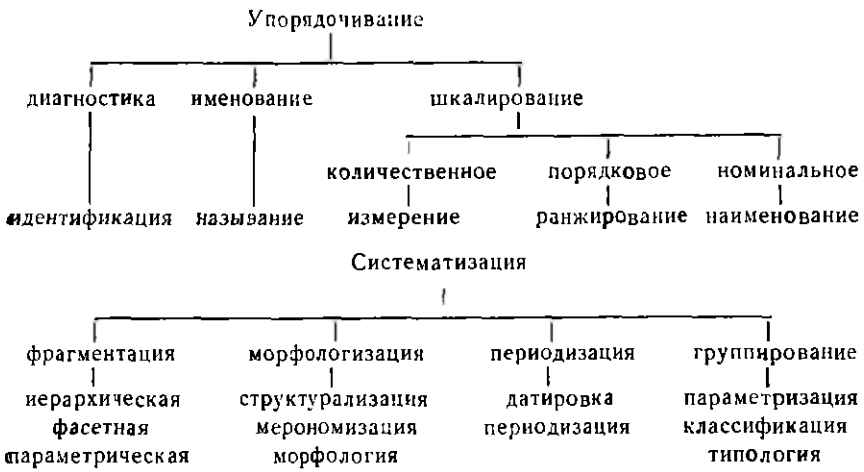
Производя подобную работу, П. М. Алексеев описывает количественные особенности устройства текста, т. е. характеризует мероны (лексические, грамматические и прочие единицы текста). При этом характеристики меронов получают через измерение количественное выражение, так что для них может быть найдено место на определенной количественной шкале. Результатом такой работы являются обобщенные образы (архетипы) текстов некой группы, представленные в виде усредненной структуры текстов определенного класса. При таком способе работы фактически конструируются типы текстов. Это своего рода типостроение, а не типология. Подобным же образом можно заниматься и квантитативным типостроением (не типологией) стилей. Построив такие типы стилей, можно заниматься типологией стилей конкретных произведений. При этом мы, по существу, перетипологизируем ранее построенные типологии. Таким образом, квантитативная типология текста является одним из средств итеративной связи таксономии и мерономии. Здесь уместно обратить внимание на то, что в науке нет проблемы собственно классификации, а есть проблема переклассификации, улучшающей предшествующую. Квантитативная типология как раз и является одним из средств такой переклассификации.

Итак, у стилеметрии есть область соприкосновения с квантитативной типологией текста, по П. М. Алексееву, в области изучения мерономических структур.

В заключение обратим внимание на то, что, определенным образом представив и проинтерпретировав материал, можно заменить один вид упорядочивающе-систематизирующей дея-

тельности другим: классификацию можно представить как *районирование в пространстве* признаков, периодизацию как районирование во времени; систему классов можно представить как номинальную шкалу и тогда атрибуцию интерпретировать как наименование и т. д. Осуществляя подобного рода операции, нужно ясно отдавать себе отчет о том, каков их *логический статус*. Сказанное в полной мере относится и к стилеметрическим задачам.

Основные виды упорядочивающе-систематизирующей работы, в том числе и те, которые относятся к компетенции стилеметрии, представлены на схеме, приведенной ниже.



Литература

Алексеев П. М. Методика *квантитативной типологии текста*. Л., 1983. 76 с.

Арапов М. В. Квантитативная лингвистика. М., 1988. 184 с.

Бородкин Л. И. 1) Математические методы классификации древних текстов // Методы качественного анализа текстов нарративных источников: Сб. статей. М., 1983. С. 8—30; 2) ЭВМ ищет авторов средневековых текстов // *Число и мысль*. Вып. 9. М., 1986. С. 113—141.

Герд А. С. Ареальная типология славянских текстов XV—XVI вв. // Сов. славяноведение. 1982. № 5. С. 74—82.

Гете И. В. Избранные сочинения по естествознанию. М., 1957. 533 с.

Головин Б. Н. Язык и статистика. М., 1971. 190 с.

Ермоленко Г. В. Анонимные произведения и их авторы. Минск, 1988. 118 с.

Кимбл Г. Как правильно пользоваться статистикой. М., 1982. 295 с.

Кокс Д., Снелл Э. Прикладная статистика: Принципы и примеры. М., 1984. 200 с.

Кэррол Дж. Б. Факторный анализ стилевых характеристик прозы // Семиотика и искусствометрия. М., 1972. С. 183—197.

Мальцева Г. Ф. Некоторые количественные приемы описания индивидуального авторского стиля // Статистика текста. Т. 1. Минск, 1969. С. 206—248.

Мартыненко Г. Я. Основы стилеметрич. Л., 1988. 174 с.

Мартыненко Г. Я., Чебанов С. В. Классификационные задачи стилеметрии // Квантитативная лингвистика и автоматический анализ текста. Тарту, 1988. С. 119—136.

Марусенко М. А. Отбор информативных параметров в задачах стилистической диагностики // Структурная и прикладная лингвистика. Вып. 3. Л., 1987. С. 84—94.

Мейен С. В. Принципы исторических реконструкций в биологии // Системность и эволюция. М., 1984. С. 84—94.

Мейн С. В., Шрейдер Ю. А. Методологические аспекты теории классификации // Вопр. философии. 1976. № 12. С. 67—79.

Методы количественного анализа текстов нарративных источников. М., 1983.

Морозов Н. А. Лингвистические спектры: Средство для сличения плагиата от истинных произведений того или иного автора: Стилеметрический этюд // Изв. Императорской Академии наук. Отд-ние яз. и словесности. 1915. XX, кн. 1—4.

Орлов Ю. К. О статистической структуре сообщений, оптимальных для человеческого восприятия // Научно-техническая информация. Сер. 2. 1970. № 2. С. 11—16.

Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика. Киев, 1981. 192 с.

Статистичні параметри стилів / Ред. В. С. Перебийніс. Київ, 1967.

Тимофеева М. К. О методике индуктивной реконструкции грамматик. Новосибирск, 1983. 20 с.

Тулдава Ю. А. Проблемы и методы квантитативно-системного анализа лексики. Таллинн, 1987. 204 с.

Фукс В. По всем правилам искусства (точные методы в исследованиях литературы, музыки и изобразительного искусства) // Искусство и ЭВМ. М., 1975.

Чебанов С. В. Внутренние и внешние системы в теории классификации // Системные исследования: Ежегодник. 1980. С. 140—146.

Чупров А. А. Очерки по теории статистики. СПб., 1910. 443 с.

Шайкевич А. Я. Гипотезы о естественных классах и возможность количественной таксономии в лингвистике // Гипотеза в современной лингвистике. М., 1980. С. 310—357.

Шрейдер Ю. А., Шаров А. А. Системы и модели. М., 1982.

Dittenberger W. Sprachliche Kriterien für die Chronologie der platonische Dialoge // Hermes (Berlin). 1880. № 16. S. 321—345.

Doyle L. B. Is Automatic Classification a Reasonable Application of Statistical Analysis of Text? // J. Assoc. Computing Machinery. 1965. Vol. 12, N 4, oct.

Kraus I., Vašak P. Попытка количественной типологии текстов // Prague Studies in Mathematical Linguistic 2. Praha, 1967. S. 77—88.

СЛОЖНОСТЬ СИНТАКСИЧЕСКИХ СТРУКТУР И СТИЛИСТИЧЕСКАЯ ДИАГНОСТИКА

ПОВЕРХНОСТНЫЙ СИНТАКСИС И ДИАГНОСТИКА

Среди форм упорядочивания, которыми занимается стилеметрия, центральное место занимает диагностика, образующая основу прикладных диагностических классификаций. Такие классификации играют особо важную роль там, где осуществляется массовая, поточная обработка данных с использованием ЭВМ и малоквалифицированного труда.

Диагностические классификации строятся на легко измеримых (в том числе инструментально) симптоматических признаках, характеризующих периферию характеристики объекта исследования (см. предыдущий раздел). Ведущую роль в стилеметрии играют, как правило, количественные признаки, не выходящие за рамки формально-грамматической, в частности поверхностной, синтаксической структуры в ее структурном, линейном, категориальном (синтаксис частей речи) и функциональном (синтаксис членов предложения) аспектах.

Центральная роль синтаксического уровня в стилеметрии определяется следующими факторами:

— индивидуальная свобода выбора структурных вариантов на синтаксическом уровне значительно выше, чем на каком бы то ни было другом уровне, за исключением, быть может, гиперсинтаксического;

— синтаксические структуры отличаются очень высокой степенью повторяемости минимальных единиц, структурных схем, связей, позиций и т. п. Это существенно упрощает организацию выборочного наблюдения и оценку достоверности полученных результатов;

— все, что с точки зрения синтаксической системы языка является несущественным, в языке художественной литературы переходит в область различения поэтических смыслов;

— в синтаксическом строе наиболее откровенно проявляется авторская манера развертывания мысли, которая может быть проинтерпретирована в «содержательных» терминах, — таких, как слитность — разорванность мысли, ее синтетичность —

аналитичность, напряженность — уравновешенность, простота — сложность, предметность — процессуальность и т. д.;

— синтаксис принадлежит к числу наиболее изученных уровней языка, а в последнее время достигнуты определенные успехи в области разработки формализованных синтаксических моделей, элементы метаязыка которых легко поддаются количественному измерению. Речь в данном случае идет о различных вариантах представления синтаксической структуры предложения в терминах теории графов (структуры составляющих, зависимостей, гибридные структуры). При таком подходе множество предложений заменяется совокупностью линейно-структурных схем, среди которых выделяется набор типовых «геометрических» фигур (мер синтаксической сложности), числовые характеристики которых выступают в роли стилистических признаков.

В высоком стилеформирующем и стилеразличающем потенциале синтаксического уровня можно убедиться, не проводя массового исследования. Для этого достаточно провести небольшой синтаксический эксперимент в духе А. М. Пешковского. Примером такого эксперимента являются стилистические вариации на одну и ту же тему, которые приводятся в книге Т. П. Крестинской и С. В. Петрова:

«Стемнело. Полил сильный дождь. Еле можно было разглядеть силуэты деревьев. Они смутно чернели в саду. Стук дождя о крышу. Откуда-то донесся рев пароходного гудка. Неожиданно. Резко».

«Когда стемнело, полил сильный дождь, и еле можно было разглядеть силуэты деревьев, которые смутно чернели в саду. Сквозь стук дождя о крышу донесся откуда-то неожиданно и резко рев пароходного гудка».

«Стемнело, и полил сильный дождь. И еле можно было раглядеть смутно черневшие в саду силуэты деревьев. И донесся откуда-то сквозь стук дождя о крышу рев пароходного гудка, и неожиданно, и резко».

«Сильный полил, когда стемнело, дождь, и еле можно было разглядеть силуэты деревьев — чернели в саду смутно. Неожиданно и резко донесся сквозь стук дождя о крышу откуда-то пароходного гудка рев» (Крестинская, Петров, 1963, с. 162).

Синтаксис первого текста — отрывистый, нервный, напряженный, напоминающий то, что называют телеграфным стилем. Такой текст можно сравнить с картиной экспрессиониста, изображающего действительность небрежно брошенными мазками.

Слог второго текста — плавный, размеренный, деловитый. Т. П. Крестинская и С. В. Петров отмечают, что такой синтаксис вполне возможен и в научной прозе, стоит лишь изменить тему и лексическое наполнение: «Когда вода подогрелась, создались оптимальные условия, и можно было рассмотреть невооруженным взглядом тельца организмов, которые быстро плавали в изменившейся среде. От внезапного и резкого стука о стену сосуда началось среди них беспорядочное, убыстрившееся движение» (там же).

Третья вариация, построенная на нагнетательном использовании союза *и*, ближе всего к сказочной или к историко-хроникальной речи.

И наконец, четвертая вариация — самая необычная, самая напряженная. Здесь и инверсия, и синтаксические разрывы, и даже то, что в теории синтаксических моделей называют непроективностью.

Число таких синтаксических вариаций можно было бы умножить, но даже из того, чем мы располагаем, довольно явственно видна ведущая роль синтаксических структур в построении «стилевого образа» каждого текста.

Состав и количество стилистических признаков, предлагаемых разными исследователями и используемыми для различных целей, варьируют в очень широких границах: от одного-двух до нескольких десятков и даже сотен. В такой неопределенной ситуации возникают естественные вопросы: Все ли признаки являются стилистически важными, и если да, то в какой степени? Сколько их должно быть? Как влияет классификационная задача на число и состав диагностических признаков? и т. п. На эти вопросы однозначного ответа пока нет. На первый взгляд может показаться, что, имея дело с такой «капризной» реальностью, как художественная речь, нужно стремиться к максимально полному набору стилистически чувствительных признаков. Такое стремление можно понять, но при всей житейской привлекательности оно не может служить основой для достаточно строгого научного исследования.

Отбор минимального числа стилистически релевантных признаков — наиболее трудный этап диагностической работы. Обычно такой отбор осуществляется с помощью процедур многомерной статистики: метода корреляционных плеяд, метода главных компонент, метода экстремальной группировки параметров и др. Все эти процедуры, однако, имеет смысл применять только тогда, когда на предэкспериментальной стадии исследователями-экспертами осуществлен качественный анализ признаков и среди них выделены те, которые затем могут претендовать на участие в диагностической классификации. Обоснованность этих претензий отвергается или подтверждается с помощью того или иного варианта многомерного анализа.

Как уже говорилось выше, в последние годы широкое распространение получило научное направление, использующее для решения диагностических задач комплекс мер синтаксической сложности, сформулированных в теории синтаксических моделей языка. Ниже приводится типология этих мер и описываются их диагностические возможности.

ИЗМЕРЕНИЕ СИНТАКСИЧЕСКОЙ СЛОЖНОСТИ

Понятие сложности в современной науке приобрело статус общесистемного параметра подобно таким понятиям, как орга-

низованность, устойчивость, целостность, разнообразие и др. В различных дисциплинах (системный анализ, кибернетика, логика науки, теория распознавания, теория коммуникации и т. п.) разрабатываются классификации разновидностей сложности и методы их количественного измерения.

Явление сложности обычно связывается с трудностями, возникающими в ходе решения различного рода задач, простота же выступает как «понятность», легкость решения, освоения, обучения и т. п. Такое понимание распространено и в лингвистике: лингвистическая сложность рассматривается как показатель успешности взаимодействия между участниками коммуникативного акта, в частности, как степень трудности порождения или восприятия речевого сообщения.

Серьезным препятствием на пути отыскания критериев оценки лингвистической сложности является множественность модусов существования языка (язык, речь, норма, узус, текст), сфер и регистров его функционирования, разнообразие семиотических планов (семантика, синтактика, прагматика), расслоение языка на отдельные уровни. Последний фактор заставляет выделять несколько видов сложности: морфологическую, синтаксическую, семантическую, сложность на уровне сверхфразовых единств, композиционную сложность и т. д. Весьма заманчивой представляется задача разыскания интегральных, комплексных характеристик сложности речи. Для того чтобы такие характеристики стали реальностью, должна быть разработана достаточно полная и научно обоснованная система оценок сложности текста в терминах каждого лингвистического уровня, установлена относительная значимость каждого вида сложности и система связей между ними.

Наибольшие успехи в области исследования лингвистической сложности достигнуты на синтаксическом уровне. Накопленные результаты уже столь значительны, что представляется своевременным упорядочить существующие представления, дать детализованную классификацию способов измерения синтаксической сложности и оценить эффективность их использования в современной лингвистике.

Для большинства лингвистов характерно расплывчатое, по существу, не выходящее за рамки повседневного опыта, понимание синтаксической сложности. К синтаксически сложным обычно относят высказывания, перенасыщенные словесным или конструктивным материалом, т. е. сложность отождествляется с многословием: «много» значит «сложно». Указывается, например, что чрезмерное насыщение цельного предложения элементарными, в особенности с разветвленной схемой сочинения и подчинения, расширенными причастными оборотами, отглагольными существительными с большим числом зависимых членов и т. п., крайне усложняет синтаксические построе-

ния, делает их тяжеловесными, громоздкими, неуклюжими, необозримыми.

Внедрение в исследовательскую практику лингвиста методов математического моделирования естественных языков значительно расширило возможности исследования механизма усложнения речи, поскольку модельный подход благодаря упрощению, «огрублению» представлений о синтаксисе высказывания дает возможность сосредоточить внимание на наиболее общих закономерностях синтаксической организации естественных языков.

Точное математическое представление синтаксической структуры предложения в терминах теории графов позволяет абстрагироваться от словесного, конструктивного и комбинаторного разнообразия элементов речевой цепи. Так, в структурах составляющих и структурах зависимостей представителями всех классов слов (грамматических и лексико-семантических) являются узлы дерева, а все типы синтаксических связей сводятся к одному — отношению подчинения в грамматике зависимостей и отношению группирования элементов, отражающему степень близости связей, в грамматике непосредственно составляющих (НС), а синтаксическая конструкция — это лишь группа узлов с определенным набором ориентированных (грамматика зависимостей) или неориентированных (грамматика НС) связей. Такой «упрощенческий» подход к структуре высказывания помимо своей геометрической общности обладает также и тем преимуществом, что он создает хорошую основу для измерения степени употребительности синтаксических фигур на массовом речевом материале.

КЛАССИФИКАЦИЯ МЕР СИНТАКСИЧЕСКОЙ СЛОЖНОСТИ

В эксплицитной, древовидной схеме предложения можно выделить три наиболее общих типа операций, используемых для усложнения структуры предложения:

— операции, основанные на усложнении структурного (иерархического) порядка единиц, образующих предложение;

— операции, основанные на усложнении их линейного порядка;

— операции, основанные на вариантах использования отношения однородности.

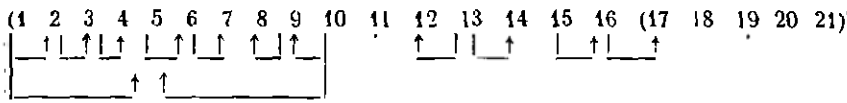
Меры сложности, основанные на операциях первого типа, мы будем называть мерами гипотактической сложности; меры, основанные на операциях второго типа, — мерами линейной сложности; меры, основанные на операциях третьего типа, — мерами паратактической сложности.

Воспользуемся гибридной схемой представления синтакси-

ческой структуры предложения, основанной на использовании стрелочной и скобочной записи в рамках соответственно теории зависимостей и теории составляющих, что призвано обеспечить различение подчинительных и сочинительных связей. Подобные гибридные способы представления синтаксической структуры использовались Л. Н. Иорданской (1967), изучались Ю. А. Шрейдером (1964), М. И. Белецким (1971), А. В. Гладким (1985).

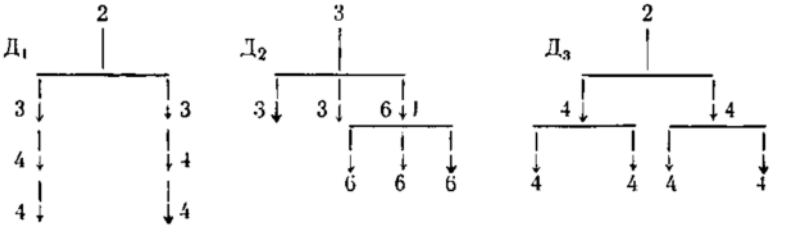
Покажем на конкретном примере используемый нами способ представления синтаксической структуры предложения.

1 2 3 4 5 6 7 8 9
«По вечерам этажи Атлантиды сияли во мраке огненными несметными
10 11 12 13 14 15 16 17 18 19
глазами и великое множество слуг работало в поварских, судомойнях и
20 21
винных подвалах» (И. Бунин. Господин из Сан-Франциско).



Среди мер гипотактической сложности можно выделить прежде всего те, которые строятся на использовании механизма параллельного и последовательного подчинения. Эти контрастные приемы в теории синтаксических моделей выступают в качестве двух мер иерархической сложности. Первую называют шириной дерева зависимостей, а вторую — длиной (или высотой) дерева. Ширина дерева в узле x равна числу стрелок, исходящих из данного узла, а длина дерева в узле x равна расстоянию от вершины дерева до данного узла, измеренному количеством стрелок. Максимальные значения ширины и длины, выделенные в каждом дереве, считаются шириной и длиной дерева в целом.

Наряду с этими двумя мерами есть еще несколько мер, которые учитывают соотношение между последовательным и параллельным подчинением. Среди этих мер можно прежде всего назвать две: густота дерева (число в нем висячих узлов) и мощность уровня (число узлов на одном уровне подчинения). Особый интерес представляет мера, предложенная Ю. А. Шрейдером (1971), ее назовем интегральной шириной. Интегральная ширина слагается из ширины дерева в данном узле и ширины узлов предыдущих ярусов, соединенных с данным узлом. Сложность дерева в целом определяется как суммарная сложность его узлов. Пусть мы имеем три семиузловых дерева:



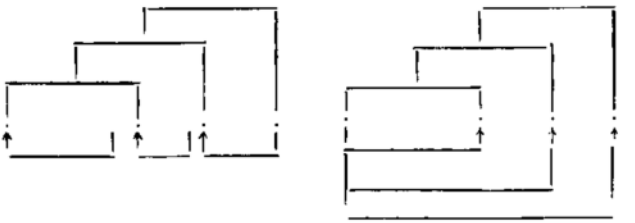
Тогда интегральная иерархическая сложность дерева $Д_1$ равна 22, дерева $Д_2$ — 33, а дерева $Д_3$ — 26. Эта мера синтаксической сложности выступает как показатель оптимального соотношения между механизмами параллельного и последовательного подчинения. Так, среди деревьев, приведенных выше, наиболее оптимальным следует считать дерево $Д_1$. В целом более оптимальны те деревья, в которых преобладает механизм последовательного подчинения с умеренным использованием техники параллельного подчинения.

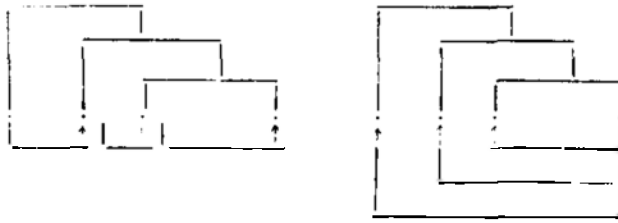
В основу классификации мер линейной сложности мы положим результаты взаимодействия вариантов иерархического порядка элементов фразы (параллельное и последовательное подчинение, смешанная зависимость) с вариантами линейного порядка (постпозиция, препозиция, интерпозиция).

В результате взаимодействия препозиции и постпозиции с цепочками последовательно зависимых слов возникают линейные последовательности с левым и правым последовательным подчинением. Количество связей в таких цепочках будем считать соответственно *левой* и *правой длиной*.

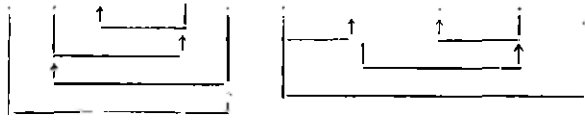
При взаимодействии постпозиции и препозиции с параллельным подчинением возникают цепочки с левым и правым соподчинением. Ширину таких цепочек будем называть соответственно *левой* и *правой шириной*.

Цепочки с левой последовательной подчиненностью и правой параллельной подчиненностью в терминах структур составляющих имеют одинаковую геометрию. То же самое можно сказать и о цепочках с правой последовательной и левой параллельной подчиненностью. Первые в теории НС называют *регрессивными* (или левоветвящимися), вторые — *прогрессивными* (или правоветвящимися).





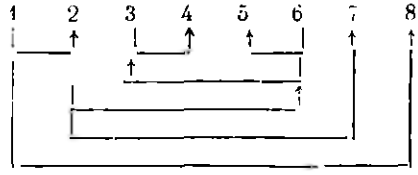
При взаимодействии интерпозиции с цепочками последовательно зависимых узлов или с цепочками со смешанной зависимостью возникают гнездующиеся конструкции:



В рамках теорий НС и зависимостей с привлечением некоторых идей из общей теории автоматов было установлено, что наиболее трудными при порождении речи являются регрессивные, а при распознавании — прогрессивные цепочки (Ингве, 1965). Что касается гнездующихся фигур, то они одинаково трудны как при производстве, так и при распознавании речи (Хомский, Миллер, 1967).

Применительно к дереву зависимостей степенью гнездования узла называют количество стрелок, накрывающих данный узел, и притом таких, которые не имеют общих краев (край — это начало или конец стрелки) (Фитиалов, 1968).

Кроме степени гнездования можно предложить еще две меры сложности, основанные на использовании приема интерпозиции: степень разрывности и степень дистантизации. Например, в дереве зависимостей



между управляющим и управляемым узлами имеется 5 разрывов (между узлами 1 и 8, 2 и 7, 2 и 6, 6 и 3, 3 и 5). Легко видеть, что степень разрывности дерева, достигаемая в узле x, равна числу стрелок, накрывающих данный узел. Степень разрывности тесно связана со степенью гнездования фразы и степенью параллельной подчиненности (как левой, так и правой): чем выше значения этих мер сложности, тем выше

степень разрывности фразы. Степенью дистантизации фразы мы будем называть наибольшее число узлов, размещенных между двумя синтаксически связанными узлами. В дереве, приведенном выше, наибольшая степень дистантизации достигается между узлами 1 и 8: между ними располагается 6 узлов.

Со степенью гнездования тесно связана еще одна мера, называемая И. П. Севбо зигзагом и измеряемая максимальным числом перемен в линейной ориентации стрелок на пути из корня дерева в висячий узел (Севбо, 1981). Так, в дереве

таких «переориентаций» четыре. Легко видеть, что эта мера сложности положительно коррелирует с высотой дерева и степенью его гнездования.

ПРОБЛЕМА МНОГОМЕРНОГО АНАЛИЗА МЕР СЛОЖНОСТИ

Меры синтаксической сложности дают комплексную характеристику предложения в его структурном и линейном аспектах, а приведенная выше типология позволяет интерпретировать эти меры с точки зрения синтаксических механизмов самого общего типа: сочинения и подчинения, постпозиции и препозиции, контактности — дистантности и т. п. Диагностические свойства мер сложности, конечно, не одинаковы: одни меры могут быть диагностически существенными, скажем, при атрибуции текстов и малосущественными при их типологии; для одного жанра одна и та же мера сложности диагностически информативна, а для другого — нет; некоторые меры сложности хотя и обладают диагностическим потенциалом, но этот потенциал — «чужой», он практически целиком дублирует информацию, содержащуюся в других мерах сложности.

Все эти неопределенности могут разрешаться различными способами. Наиболее очевидный — содержательный, он основан на компетенции или интуиции исследователя, выступающего в качестве эксперта. Второй способ — формальный. Он основывается на применении той или иной алгоритмической процедуры многомерного анализа данных: факторного анализа, метода корреляционных плеяд, метода минимального корреляционного пути и т. п. В эпоху тотальной компьютеризации второй путь многим исследователям представляется более пер-

спективным, он их увлекает тем, что метод сам по себе, независимо от исследователя, позволяет обнаруживать скрытые закономерности в огромной массе эмпирических данных. Однако значительная часть исследователей к вере в безграничные возможности метода относятся весьма скептически, и не только потому, что она покоится на пороках наивной философии эмпиризма, а прежде всего потому, что в таком подходе четко просматриваются несколько методических изъянов:

— результаты многомерного анализа, как правило, не сопровождаются статистическими заключениями о значимости различий между эмпирическими и теоретическими величинами; это в значительной мере затрудняет проверку надежности применяемых процедур, не выходящих, несмотря на сложность и изощренность математического аппарата, за рамки описательности, присущей традиционным статистическим методам;

— большинство методов многомерного анализа исходят из весьма сильного допущения о нормальности исследуемых совокупностей, а встречи с нормальным законом в лингвистических измерениях могут быть только случайными;

— все процедуры многомерного анализа признаков основаны на применении коэффициента корреляции. Среди множества связей между двумя переменными коэффициент корреляции учитывает только одну — линейную связь. Различные виды нелинейных связей (а в тексте такие связи доминируют) могут иметь величину коэффициента корреляции, подозрительно близкую к нулю, если интерпретировать ее без обращения к диаграмме рассеивания. Если между переменными x и y существует тесная линейная связь, то смысл коэффициента корреляции совершенно ясен. Однако если x и y связаны какой-то нелинейной связью, то близкие к нулю значения коэффициента корреляции могут быть получены даже несмотря на то, что между x и y существует тесная связь. Возможна и обратная ситуация: x и y могут быть связаны сильной линейной связью, но коэффициент корреляции вопреки этому может быть весьма небольшим. Это наблюдается в тех случаях, когда распределение частот обеих переменных скошены в разные стороны. Если учесть, что коэффициент корреляции рассчитывается исходя из предположения, что распределения коррелирующих признаков описываются двумерной нормальной поверхностью, то правомерность его использования становится еще более проблематичной.

В сказанном можно отчасти убедиться, обратившись к результатам регрессионного анализа мер сложности.

РЕГРЕССИОННЫЙ АНАЛИЗ МЕР СЛОЖНОСТИ

Характер и тесноту корреляционной связи между мерами сложности исследуем на материале русской художественной прозы конца XIX — начала XX в. Рассмотрим тексты десяти

авторов с резко различной индивидуальной манерой письма — Л. Андреева, А. Куприна, В. Вересаева, И. Бунина, Ф. Сологуба, Л. Толстого, Д. Мамина-Сибиряка, А. Чехова, В. Дорошевича, Д. Мережковского.

Ограничимся анализом самых простых зависимостей.

Приступая к статистическому анализу, нужно иметь в виду, что диагностические признаки могут быть экстенсивными (объемными) — размеры словоформы, предложения, абзаца и т. п. — и интенсивными (структурными). Между этими двумя группами признаков имеется довольно тесная статистическая связь. Например, числовые значения конкретных мер сложности в сильной степени зависят от склонности автора к употреблению коротких или длинных предложений. Естественно ожидать, что чем больше будет у данного автора средний размер предложения, тем большим будет среднее значение всех других параметров. Поэтому разумно снять «возмущающее» воздействие размера предложения и сосредоточить внимание исключительно на внутренне присущих данному автору структурных механизмах. Если этого не сделать, то некоторые существенные особенности индивидуальной манеры письма могут ускользнуть из поля зрения исследователя, поскольку размер предложения от автора к автору варьирует в очень широких границах.

Для снятия возмущающего воздействия размера предложения была построена серия корреляционных таблиц, отражающих зависимость между размером предложения и остальными параметрами. При обработке количественных данных были получены два ряда условных средних в соответствии с тем, какой из двух признаков (размер предложения или величина параметра) выступал в роли факторного или результативного. Здесь мы ограничимся рассмотрением варианта, когда в роли факторного признака выступает размер предложения, а в роли результативного — значения шести интересующих нас параметров.

Остановимся сначала на зависимости гипотактических параметров (ширины w и высоты l дерева) от размера предложения. На рис. 1 и 2 видно, что условные средние этих параметров постепенно замедляют свой рост. (На рис. 2 обозначены: *1* — Д. Мережковский, *2* — В. Дорошевич, *3* — Л. Толстой, *4* — Ф. Сологуб, *5* — Л. Андреев.) Особенно быстро это происходит с условными средними параметра w: начиная с размера предложения, равного примерно 35 словам, регрессионная кривая практически достигает уровня насыщения и стелется вдоль оси абсцисс. Из этого следует, что две разновидности гипотаксиса ведут себя принципиально иначе: оба параметра неограниченно возрастают, но теоретически у w верхний предел есть, а у l такого предела нет.

Однако различия между двумя гипотактическими парамет-

рами этим не исчерпываются. Ширина дерева обладает тем замечательным свойством, что ее средняя величина в предложениях любой фиксированной длины не зависит от индивидуальной манеры письма, а лишь удостоверяет принадлежность

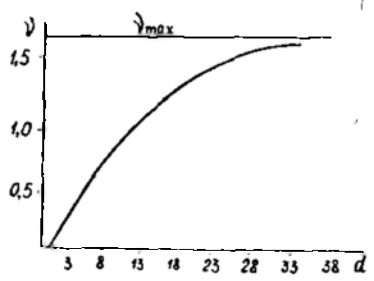


Рис. 1.

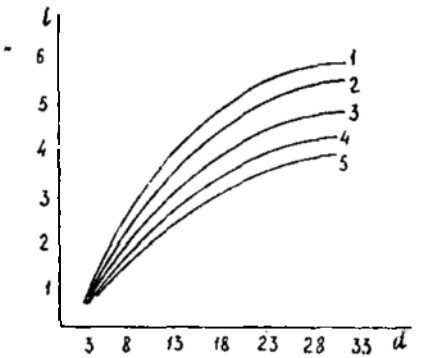


Рис. 2.

текстов конкретного автора к жанру художественной прозы. Ширина дерева от автора к автору, конечно, варьирует, но эта вариация имеет отраженный характер: она целиком определяется варьированием среднего размера предложения. У авторов с одинаковым средним размером предложения w также будет одинаковой.

Асимптотический характер возрастания условных средних w с хорошим приближением описывается экспоненциальной функцией вида

$$w = w_{\max}(1 - e^{-c(d-1)}),$$

где $w_{\max}$ — уровень насыщения, d — размер предложения, c — постоянный коэффициент.

Величина уровня насыщения $w_{\max}$ не зависит от принадлежности текстов к тому или иному функциональному стилю, а определяется исключительно структурой конкретного естественного языка. Иначе говоря, этот коэффициент можно считать внутриязыковой универсалией: в русском языке он равен 3,8 соподчиняющих связей. В отличие от $w_{\max}$ коэффициент c имеет фиксированное значение лишь в пределах одного функционального стиля или жанра в пределах этого стиля, т. е. является внутристилевой (внутрижанровой) универсалией: для русской художественной прозы исследуемого периода он равен 0,10.

Что касается высоты дерева l, то в рамках художественной прозы каждый автор имеет индивидуальную регрессион-

ную кривую в виде восходящей ветви параболы с предельно простым аналитическим выражением:

$$l = (d - 1)^n, \qquad (1)$$

где n — коэффициент, имеющий у каждого автора фиксированное значение.

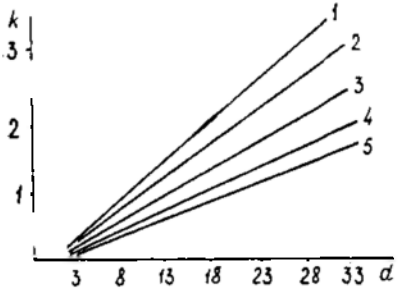


Рис. 3.

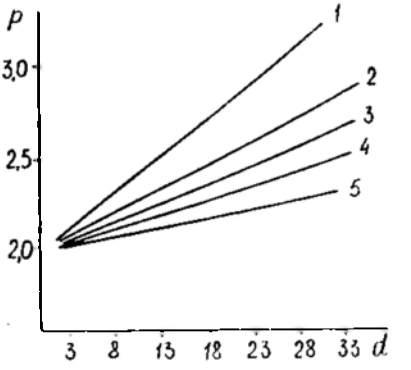


Рис. 4.

Паратактические параметры (степень однородности k и степень перечислительности p) ведут себя принципиально иначе. В отличие от гипотактических параметров их условные средние возрастают с постоянной скоростью, т. е. подчиняются линейному закону. При этом каждый автор имеет индивидуальную прямую регрессии (рис. 3: *1* — Л. Андреев, *2* — Ф. Сологуб, *3* — Л. Толстой, *4* — В. Дорошевич, *5* — Д. Мережковский; рис. 4: *1* — В. Дорошевич, *2* — Д. Мережковский, *3* — Ф. Сологуб, *4* — Л. Толстой, *5* — Л. Андреев). Из этого следует, что параметры k и p выполняют стилеразличающие функции. Аналогичная закономерность выявлена и в поведении степени дистантности s (рис. 5: *1* — В. Дорошевич, *2* — Ф. Сологуб, *3* — Д. Мережковский, *4* — Л. Толстой, *5* — Л. Андреев). Регрессивные уравнения для этих трех параметров также очень просты.

$$k = a(d - 1), \qquad (2)$$

$$p = b(d_k - 2) + 2, \qquad (3)$$

$$s = f(d - 2), \qquad (4)$$

где d_k — размер предложения с однородностью, a, b, f — коэффициенты, имеющие для каждого автора фиксированное значение.

При $d^k = 2$ правая часть уравнения (3) равна 2, т. е. минимально возможному объему однородной группы.

Что касается степени гнездования, то в поведении этого параметра и поведения ширины дерева наблюдается полный параллелизм. Возрастание условных средних ν также имеет асимптотический характер, и регрессионные кривые подчиняются тому же закону, что и параметр w. При этом различия между индивидуальными кривыми регрессии статистически не значимы: они практически сливаются в одну линию (рис. 6). Это означает, что степень гнездования, как и степень соподчиненности, является внутрижанровой универсалией.

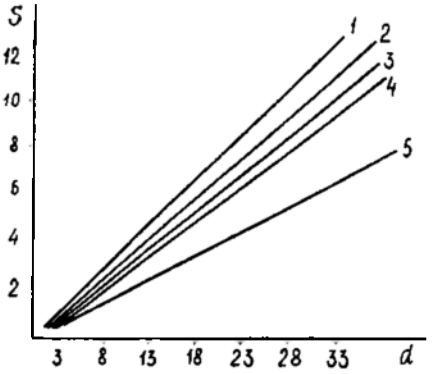


Рис. 5.

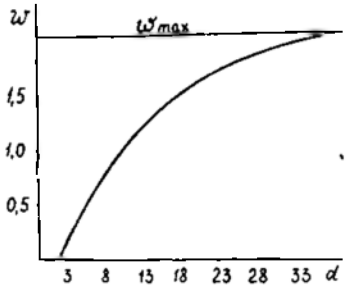


Рис. 6.

Пучки линий регрессии, показанные на рис. 1 — 4, дают наглядное представление о синтаксисе художественной прозы исследуемого периода. По размаху веера каждого пучка можно судить, например, о верхней и нижней границах степени паратактичности или гипотактичности беллетристики, а также о том, в прозведениях каких писателей эти пределы достигаются. Из графиков, в частности, видно, что предложение Л. Андреева превосходит предложения других писателей по степени однородности, а по высоте дерева, наоборот, им уступает. Полярная противоположность Л. Андрееву по этим параметрам — Д. Мережковский. Остальные восемь прозаиков группируются между двумя синтаксическими «антиподами», тяготея в той или иной мере к крайним значениям паратаксиса или гипотаксиса или стремясь уравновесить две противоборствующие тенденции. Конкуренция между этими тенденциями отличается удивительной правильностью. На рис. 2 и 3 видно, что индивидуальные линии регрессии параметров k и l между осями координат располагаются в строго обратном порядке. Если построить два ранжированных ряда (убывающий для степени однородности и возрастающий для степени подчиненности или наоборот), воспользовавшись для этого индивидуальными значениями коэффициентов a и n в регрессионных уравнениях (1), (2), то окажется, что в

двух ранжировках нет заметных расхождений. Статистически это подтверждается тем, что коэффициент ранговой корреляции в данном случае практически не отличается от единицы. Следовательно, при изучении внутрижанровой дифференциации текстов одним из этих параметров можно пренебречь. Поскольку размах веера регрессионных кривых у степени однородности значительно шире, этот параметр имеет больше оснований для участия в диагностических классификациях.

Из проведенного анализа можно сделать следующие предварительные выводы.

1. Связи между мерами сложности, как, по-видимому, и другими диагностическими признаками, часто имеют нелинейный характер, причем нелинейные связи неоднозначны, некоторые из них характеризуются асимптотическим ростом, другие же возрастают монотонно.

2. В статистическом смысле нелинейная связь может быть достаточно сильной вопреки малой величине коэффициента корреляции, занимающего силу реально существующей связи. В принципе, возможна обратная ситуация.

3. Стилеразличающие возможности некоторых диагностических признаков (например, ширины дерева и степени гнездования) целиком определяются воздействием экстенсивных параметров, например размера предложения. После снятия этого возмущающего фактора они превращаются во внутристилевые или внутрижанровые константы.

4. Связь между некоторыми диагностическими признаками после снятия возмущающего воздействия размера предложения приближается к функциональной (почти строго прямо или обратно пропорциональной).

Описанные результаты не имеют абсолютного характера, это, скорее, осторожный набросок некоторых тенденций, которые наблюдаются в реальных текстах. Теория и методика стилистической диагностики находятся пока в начальной стадии своего становления.

Литература

Белецкий М. М. Математическая модель предложения с однородностью // Научно-техническая информация. Сер. 2. 1971. № 4. С. 30—36.

Вопросы статистической стилистики. Киев, 1974.

Гладкий А. В. Синтаксические структуры естественного языка в автоматизированных системах общения. М., 1985.

Долинина И. Б. Системный анализ предложения. Л., 1977.

Ингве В. Гипотеза глубины // Новое в лингвистике. Вып. 4. М., 1965. С. 126—138.

Иорданская Л. Н. Автоматический синтаксический анализ. Т. 2. Новосибирск, 1967.

Крестинская Т. П., Петров С. В. Как пользоваться словом. Л., 1963.

Мартыненко Г. Я. 1) Статистическое исследование синтаксической сложности предложения (на материале болгарского языка) // Информаци-

онные вопросы семиотики, лингвистики и автоматического перевода. Вып. 1. М., 1971. С. 84—101; 2) Основы стилеметрии. Л., 1988.

Марусенко М. А. Отбор информативных параметров в задачах стилистической диагностики // Структурная и прикладная лингвистика. Л., 1987. С. 84—94.

Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика. Киев, 1981. С. 192.

Фитиалов С. Я. Об эквивалентности грамматик НС и зависимостей // Проблемы структурной лингвистики 1987. М., 1968. С. 72—102.

Хомский Н., Миллер Дж. Конечные модели пользования языком // Кибернетический сборник. Нов. сер. Вып. 4. М., 1967. С. 141—218.

Шрейдер Ю. А. 1) Свойство проективности языка //Научно-техническая информация. 1964. № 8. С. 38—46; 2) Равенство, сходство, порядок. М., 1971.

КОМПЬЮТЕРНЫЕ АСПЕКТЫ СТИЛЕМЕТРИИ

АВТОМАТИЗАЦИЯ СТИЛЕМЕТРИЧЕСКИХ ИССЛЕДОВАНИЙ

Филологические задачи, при решении которых используются сведения о языке, имеют, как правило, четкую прикладную направленность, язык и стиль текста в такой ситуации является не целью исследования, а средством решения внелингвистических проблем. Важным является также и то, что решение конкретной филологической задачи (например, проблемы спорного авторства) обычно не ограничивается жесткими рамками какой-либо одной методики исследования, а ведется с использованием методов и фактов, относящихся к самым разным областям знания и практической деятельности. Филологический анализ всегда требует интегрального, комплексного подхода. Это предполагает и внутритекстовой анализ, и анализ условий создания текста, и изучение взаимоотношений данного текста с другими текстами. Заключения филолога часто субъективны, ибо строятся на здравом смысле, догадке, интуиции, житейской мудрости и редко претендуют на безусловную научную строгость и окончательность выводов.

В стилеметрии — как и в традиционной филологии — текст остается исходной эмпирической данностью, однако стилеметрический интерес концентрируется прежде всего на количественной организации текста, а позиция исследователя характеризуется объективностью и методической строгостью (Мартыненко, 1988). В компьютерной среде эти познавательные принципы реализуются с предельной жесткостью, приобретают характер категорического императива. Видение текста здесь становится «вещным», процедурно-технологическим, исследователь с помощью компьютерных программ общается с текстом, как с непосредственно «осязаемым» материальным образованием, построенным по конкретным лингвополиграфическим законам. В компьютерной стилеметрии в существенной мере утрачивается комплексность традиционного филологического подхода, теряет-

ся бесконечное богатство ассоциаций, возникающих при общении с текстом и его окружением. И в то же время приобретаются практически неограниченные возможности для единообразной и быстрой обработки больших массивов данных, извлекаемых из печатного текста. Эффективность компьютерной работы с текстами особенно велика там, где используются сложные процедуры многомерного анализа текста: дистрибутивно-статистический метод, алгоритмы лингвистической дешифровки, методы квантитативной типологии и таксономии текстов, алгоритмы стилистической диагностики.

ОСНОВНЫЕ ЗАДАЧИ КОМПЬЮТЕРНОЙ СТИЛЕМЕТРИИ

В области компьютерной стилеметрии можно выделить следующие основные направления исследований.

1. Теоретические исследования: изучение количественных закономерностей в символьных последовательностях, исследование условий действия закона больших чисел в тексте, разыскание робастых лингвостатистических методов, устойчивых статистик и т. п.

2. Первичная обработка лингвистических данных: построение рядов распределения, вычисление статистик, статистическое оценивание, проверка статистических гипотез, построение теоретических моделей по опытным данным.

3. Систематико-таксономические задачи:

А. Обработка многомерных данных с использованием стандартных алгоритмических процедур: факторного, дискриминантного, кластерного и других методов многомерного анализа.

Б. Обработка лингвистических данных с помощью специальных лингвистических методов: дешифровочных алгоритмов, дистрибутивно-статистического метода, методов датировки, атрибуции, диагностики и типологии текстов.

4. Лексикографическая обработка текстовых данных: создание частотных и алфавитно-частотных словарей, словарей-конкордансов, словоуказателей, словарей морфем, рифм, словарей писателей, словарей-минимумов, словарей ключевых слов, ассоциативных словарей и т. п.

5. Информационно-поисковые задачи:

А. Автоматический поиск текстов (авторский, жанровый, историко-хронологический, библиографический и т. п.).

Б. Поиск текстовых единиц, обладающих определенным набором качественных и количественных характеристик для решения стилистических и грамматических проблем.

6. Лингводидактические задачи:

А. Словарное обеспечение преподавания русского языка, включая разработку словарей-минимумов, двуязычных словарей, обратных словарей и т. п.

Б. Разработка систем программированного обучения с использованием информации о жанрово-стилевой дифференциации текстов.

ПРИНЦИПЫ СТРУКТУРАЛИЗАЦИИ ТЕКСТОВ

Все сложные алгоритмические процедуры автоматической обработки текста являются разновидностями диагностической работы, связанной с упорядочиванием и систематизацией текстов, частей текста и единиц текста. Важным этапом такой работы является выделение единиц текста, его членение. Ориентация на компьютерную обработку диктует необходимость обращения к объемно-композиционному расположению речевого материала в тексте, эксплицитно выраженному в правилах и нормах графического представления и пространственного размещения единиц и частей текста. В лингвополиграфической структуре текста помимо чисто внешних примет (строкоделение, шрифт, формат, длина абзацного отступа, способы оформления заголовков и т. п.) преломляются стилистические (функциональные, жанровые, авторские и др.) особенности текста.

Компьютерную экспликацию лингвополиграфического членения текста будем в дальнейшем называть структуризацией, подчеркивая тем самым компьютерный, формальный аспект процесса членения текста. Структуризация основывается на трех принципах: формальности, однородности и лингвостилистической надежности.

Принцип формальности означает, что процесс членения текста опирается на элементы графического оформления текста с учетом правил русской пунктуации и пространственного расположения материала на страницах печатных изданий. Использование словарной информации допускается лишь как вспомогательное средство, например, при анализе стандартных сокращений, анализе сочетаний, эквивалентных слову, разрешения омонимиии «дефис — перенос».

Принцип однородности членения предполагает автоматическое отнесение единиц текста к одному из структурно-функциональных типов речи в пределах конкретного жанра. Например, в пределах художественной прозы автоматически выделяются и группируются подмножества единиц, относящиеся к разным типам письменной речи (речи автора, чужой речи) с последующим разбиением подмножества единиц чужой речи на три однородные совокупности: прямую (диалогическую) речь, смешанную речь и вложенную прямую речь. Принцип однородности членения по своему содержанию близок к контактно-вариативному членению текста, но отличается от него типологией членения, поскольку базируется на общих пунктуационно-графических особенностях оформления чужой и авторской речи в печатных (полиграфических) текстах. В жанре художественной

прозы этот принцип распространяется на такие единицы, как словоупотребление, предложение, абзац; он направлен на обеспечение достоверности лингвостатистических и стилеметрических исследований.

Принцип лингвостилистической надежности предполагает отказ от гарантированного результата как итога работы абсолютно безошибочного алгоритма членения текста. При необходимости ошибки структуризации устраняются филологом-исследователем в диалоге с компьютерной системой. Сама же лингвостилистическая надежность членения зависит от функционально-стилистической и жанровой принадлежности конкретного текста, так же как и от индивидуальной авторской манеры письма.

Практическая целесообразность структуризации текста хорошо прослеживается на примере создания машинного фонда лексических единиц. Применяемое ныне правило определения контекста как определенного числа символов вправо и влево от ключевого слова не может считаться безупречным и используется только потому, что текст, хранимый в памяти компьютера для получения карточек-цитат, не структурирован. При наличии структурно-фрагментированного текста понятие «контекст» имеет неформальное (смысловое) значение даже в формальных (компьютерных) системах, а сам контекст по желанию исследователя может строиться из произвольного числа единиц различных уровней: словоупотреблений, предложений, абзацев.

Лингвополиграфический подход к членению текстов и разработанный на его основе формально-пунктуационный метод структуризации получил свое практическое воплощение в компьютерной диалоговой системе DISSKOTE (автор О. Н. Гринбаум), результатом работы которой являются массивы структурированных текстов, образующих часть лингвистической базы данных системы ЛИНДА, которая создается на кафедре математической лингвистики СПбГУ под руководством Г. Я. Мартыненко.

ФОРМАЛЬНО-ПУНКТУАЦИОННЫЙ МЕТОД СТРУКТУРАЛИЗАЦИИ ХУДОЖЕСТВЕННОЙ ПРОЗЫ

Формально-пунктуационный метод структуризации основан на детальном анализе правил русской пунктуации и трансформации этих правил в формально-логическую схему распознающего автомата — алгоритмическое ядро диалоговой системы DISSKOTE.

Минимальное требование к введенному в ЭВМ тексту — выделенность прописных букв, использование которых является одним из основных, принципиальных моментов в русской орфографии. В современном русском языке прописные буквы употребляются для выделения: а) нового отрезка текста (глав, аб-

зацев, предложений и т. п.) — и в этом проявляется их важная роль сигнализатора членения текста; б) различных типов лексических единиц (имена собственные и квазисобственные, буквенные аббревиатуры и др.) — в этих случаях прописные буквы выполняют свою семантическую (идеографическую) функцию.

Разработка формально-пунктуационного метода структуризации художественной прозы натолкнулась на ряд существенных противоречий между нечеткостью (размытостью) пунктуационной системы, с одной стороны, и строгостью формализованной записи алгоритма — с другой. Нечеткость использования знаков препинания, отсутствие обязательной регламентации и почти стихийный процесс исторического развития пунктуации — все это, видимо, тормозило развитие работ по формальной структуризации текстов, тем более что художественные тексты, увидевшие свет в разное время (например, с прошлого века и по сегодняшний день), несут в себе не только характерные для своей эпохи пунктуационные особенности, авторские предпочтения и вольность при употреблении тех или иных знаков, но и редакторскую правку последующих изданий.

В качестве примера рассмотрим вопрос употребления тире как дополнительного знака после разграничительных знаков препинания (точки, вопросительного и восклицательного знаков, многоточия). Современными правилами пунктуации употребление тире в начале предложения — после разграничительного знака в конце предыдущего предложения и за исключением их (тире) использования в диалогической речи не предусмотрено; более того, этот вариант и ранее к общеупотребимым не относился. Тем не менее в произведениях писателей XIX в. мы встречаем:

Над каждой крышей высокий шест скворечницы; над каждым крылечком вырезной железный крутогривый конек. — Неровные стекла окон отливают цветами радуги.

(И. С. Тургенев. Деревня)

Причины такого авторского употребления тире кроются, конечно же, в стремлении ярче, отчетливей выразить границу смысла высказывания, дополнительными средствами очертить их стилистическую окраску, новую тематику, смену фактов или смену чувств и т. п. Такие случаи, в отличие, скажем, от использования тире после тех же разграничительных знаков в диалогической речи, следует отнести к особенностям индивидуальной манеры письма:

— Бутылку пива! — Значит — так нельзя? Любопытно.

(А. М. Горький. Жизнь ненужного человека)

Эти случаи не только питают наши размышления о стилистических авторско-индивидуальных предпочтениях и жанрово-

исторических особенностях, но и требуют — учитывая проблему членения текста — своей систематизации и формализации.

Выделение предложения — основной единицы текста при структуризации — построено на анализе правого и левого окружения знаков препинания, потенциально выступающих в качестве разграничителей предложений: точки, вопросительного и восклицательного знаков, многоточия. Анализ окружения таких знаков требуется потому, что в художественных текстах знаки «.», «!», «?» помимо своего традиционного места в конце предложения часто встречаются и в середине предложения, тогда как для других типов текстов это практически исключено.

Формально-пунктуационный подход диктует нам и необходимость введения операционального определения предложения. Предложением (письменного варианта текста) называется последовательность символьных цепочек и пунктуационных знаков между ними от одного конечного знака до другого. Здесь «символьная цепочка» — это текстовое графическое изображение словоупотребления, а «конечный знак» представляет собой составной пунктуационно-пространственный отрезок текста, позволяющий формально распознавать в строке ситуацию «конец предложения». Например, отрезок текста ⟨точка+тире+прописная буква⟩ является «конечным знаком» в вышеприведенном фрагменте из рассказа И. С. Тургенева «Деревня».

Общая схема принятия решения относительно выделения из текста очередного предложения выглядит — в самом общем виде — следующим образом:

а) разпознается символ, принадлежащий множеству потенциальных разграничителей предложений;

б) производится анализ правого или левого окружения выделенного символа (поиск и фиксация «конечного знака»);

в) предложение выделяется лишь в том случае, если поиск завершается в терминальной вершине «конечный знак»; в противном случае процедура повторяется для следующего разграничительного знака препинания.

Разработка машинного алгоритма выделения предложений потребовала тщательного анализа и систематизации известных правил пунктуационного оформления предложений. Целью такого исследования было составление перечня всех возможных и встречающихся в реальных художественных текстах ситуаций, в которых символы окончания предложений (знаки «.», «!», «?») действительно означают конец предложения, а также тех ситуаций, в которых это не так. В результате выявлены три варианта ситуаций, безусловно фиксирующих конец предложения:

А) ZAG-ситуация: знак перед заголовком или подзаголовком;

Б) AO-ситуация: знак перед началом нового абзаца (перед появлением в тексте абзацного отступа);

В) FIN-ситуация: знак внутри абзаца в окружении других

знаков или символов, образующих вместе с ним решающую ситуацию.

Вариант А очевиден: заголовок (подзаголовок) не может разрывать предложение на части, следовательно, умение формально выявлять в тексте заголовки (подзаголовки) позволит безошибочно фиксировать конец предшествующих им предложений.

Вариант Б тоже достаточно прозрачен, но требует некоторой детализации. Дело в том, что предложения, состоящие из *прямой речи и предшествующих ей вводящих слов автора*, нередко размещаются на разных строчках печатного текста; в таких случаях вводящие слова автора заканчиваются двоеточием, а прямая речь оформляется абзацным отступом (АО), за которым следует тире. Например:

Она вдруг почувствовала себя оскорбленной и сказала холодно:
— Нам нужно расстаться на некоторое время, а то от скуки мы можем...

(А. П. Чехов. Попрыгунья)

Наличие АО, таким образом, сигнализирует о конце предложения лишь в том случае, если новая строка (включающая в себя этот АО) не начинается с тире. Если же после АО следует тире, а предыдущая строка заканчивается двоеточием, то АО-ситуация ⟨двоеточие+АО+тире+ПБ⟩, где ПБ— прописная буква, решающей не является. Если двоеточия в конце строки, предшествующей АО, нет, то ситуация ⟨АО+тире+ПБ⟩ является решающей.

Реже в художественно-прозаических текстах встречается такой вариант, когда после АО используются кавычки, и еще реже — и тире, и кавычки; поэтому в итоге решающей оказывается АО-ситуация более общего вида, а именно: ⟨— (двоеточие) + АО⟩, где «—» — знак логического отрицания.

Вариант В является наиболее сложным, поскольку выявление FIN-ситуации производится внутри абзаца и, следовательно, может базироваться только лишь на анализе взаимного расположения знаков и символов вокруг потенциального центра решающей ситуации. Эти центры, как мы уже отмечали, образуются знаками «.», «?» и «!». Обозначим множество таких знаков препинания через Z, тогда: Z={z1, z2, z3} и z1=«.», z2=«?», z3=«!».

Любая FIN-ситуация рассматривается как последовательность трех более простых ситуаций F, I и N, т. е. FIN=(F+I+ +N). Рассмотрим эти ситуации более подробно.

1. F-ситуация (конец предложения) фиксируется при наличии одного из элементов множества F=Z+T, где T={пусто, кавычки, закрывающая скобка}. Иначе: F={Z, Z», Z)}, что соответствует любому набору символов, существенному для формального выявления в тексте ситуации «конец предложения».

2. I-ситуация (интервал) определяется теми символами, ко-

торые используются для заполнения интервала между двумя предложениями, расположенными внутри одного абзаца: I= ={i1, i2} и i1=пробел, i2=тире.

3. N-ситуация (начало предложения) фиксируется при наличии одного из элементов множества N=P+E, где P={пусто, кавычки, открывающая скобка}, а E — множество всех прописных букв. Иначе: N={E, «E, (E}, что соответствует набору символов, используемому для графического изображения начала любого предложения, находящегося в середине абзаца.

Приведем два примера, поясняющие введенные формализмы.

Придет, бывало, к нам в деревню; я-то ему племянником довожусь. — «Что, брат, Вася, — скажет, — приди-ка, брат, переночуй-ка у меня!»

(И. С. Тургенев. Три встречи)

Здесь FIN-ситуация формируется следующим образом: F= =«точка», I=i2-«тире», P=«кавычки», E=«Ч». Следовательно, FIN=⟨точка=тире+кавычки+Ч⟩.

Да ко всему-то впридачу, кроме позора-то, ненавистную жену ввести в дом! (Потому что ведь ты меня ненавидишь, я это знаю!)

(Ф. М. Достоевский. Идиот)

В этом примере FIN=⟨!+пробел+(+П⟩.

В табл. 1 формализованно представлены правила поиска ситуации «конец предложения», где по вертикали расположены ситуации F, I и N, а по горизонтали — их возможные значения.

Таблица 1. **Правила выявления ситуации «конец предложения» в середине абзаца**

F-ситуация <конец предложения>		Z			Z»			Z)		
I-ситуация <интервал>	пробел	+	+	+	+	+	—	+	+	—
	тире	+	—	—	+	+	—	+	—	—
N-ситуация <начало предложения>		E	«E	(E	E	«E	(E	E	«E	(E

В табл. 1 зафиксировано 18 возможных значений FIN-ситуаций. Знаком «+» отмечены такие последовательности символов, которые, реально встречаясь в художественных текстах, позволяют с уверенностью распознавать ситуацию «конец предложения» внутри любого абзаца текста. Здесь же знаком «—» отмечены те ситуации, которые в реальных текстах практически не встречаются; они, тем не менее, также относятся к решающим FIN-ситуациям.

Потенциальные возможности графических средств оформле-

ния конца — начала предложения оказываются, как мы видим, комбинаторно избыточными: только 70% этих возможностей реально используются в графике русского языка. В более общем плане следует сказать, что пунктуационная комбинаторная избыточность графических средств русского языка, выявленная при разработке компьютерных правил структуризации, представляет собой еще один пример избыточности выразительных средств языка в целом.

Подведем некоторые итоги. Рассматривая варианты фиксации в тексте ситуации «конец предложения» и имея в своем распоряжении только изобразительные средства письменной речи, т. е. буквы алфавита и знаки препинания, мы выделили три возможных случая: ситуация «конец предложения» фиксируется перед заголовком (ZAG-ситуация), перед началом абзаца (AO-ситуация) и внутри абзаца (FIN-ситуация). Каждый из этих случаев может быть формализован: надо лишь обеспечить поиск одновременно в двух смежных строчках анализируемого текста, сначала во внешнем цикле (анализ для текущей строки), а затем — если выявлен потенциальный центр решающей ситуации — и поиск внутренний (либо в той же, либо в следующей строке текста).

Анализ пунктуационной системы русского языка и приемов индивидуальной пунктуации, разработки компьютерных правил формального выделения единиц текста на основе опорных признаков позволили уточнить традиционное представление о предложении и дать ему новое операциональное толкование. Реализация формально-пунктуационного метода членения текстов в диалоговой компьютерной системе DISSKOTE и анализ структуризации текстов художественной прозы М. Ю. Лермонтова, А. П. Чехова, М. А. Шолохова, А. С. Серафимовича и других писателей подтвердили правильность выбранного подхода к решению проблемы выделения единиц текста и высокую лингвистическую надежность метода структуризации.

РАССЛОЕНИЕ ТЕКСТОВ НА КАЧЕСТВЕННО ОДНОРОДНЫЕ ФРАГМЕНТЫ

Однородность изучаемых объектов — важное требование лингвостатистических и стилеметрических методов исследования текстов. Этот принцип приобретает центральное значение при изучении художественных произведений, поскольку художественность предполагает прежде всего разноплановость, образность, широту использования изобразительно-выразительных языковых средств.

Своеобразие стиля художественного произведения проявляется каждый раз в том конкретном «образе автора» или «образе рассказчика», который формируется единством конкретных лексических, синтаксических, интонационных и других средств языка. Именно речевой «образ автора» определяет весь тон

повествования, особенность индивидуально-авторского стиля художественного произведения.

Сочетание двух принципов изображения действительности — извне и изнутри, с позиции автора и персонажей — реализуется на письме сложным переплетением и пространственным взаимодействием двух типов письменной речи: авторской и чужой. Не вызывает сомнений тот факт, что авторская речь является основой речевого «образа автора», следовательно, если рассматривать проблему однородности единиц художественного текста не вообще, а относительно какого-либо определенного признака, то сам выбор такого признака очевиден — это оппозиция «речь автора/чужая речь».

Лексические, синтаксические, стилистические различия, и нередко резкие различия, между авторской и чужой речью приводят, например, к тому, что прямая речь плохо поддается анализу с помощью существующих моделей языка. В то же время сравнение прямой речи в разных произведениях и ее сопоставление с разговорной речью в живом общении может дать весьма информативные результаты при изучении влияния индивидуальной авторской манеры письма на форму содержания.

Итак, стилистическому исследованию художественной прозы должно предшествовать разграничение, расслоение речевого материала на две категории — авторскую речь и чужую речь. Чужая речь включает в себя прямую речь, диалоги, прямую речь внутри прямой речи и другие виды неавторского повествования. Имея в своем распоряжении только лишь формальные средства — знаки пунктуации и символы графической системы языка, мы предприняли попытку найти такие изобразительные средства, которые позволили бы автоматически (и на базе метода структуризации) вести разграничение единиц текста по типам письменной речи. Наш алгоритм расслоения основан на строгом описании средств оформления прямой речи и ее пространственном расположении на страницах печатных изданий. Основными единицами текста при этом являются абзац и предложение, что же касается словоупотребления, то они входят в состав предложений и могут участвовать в исследованиях как единицы лексического уровня авторского или неавторского повествования.

В алгоритме расслоения основным структурным элементом текста считается абзац: все предложения одного абзаца, последовательно вычленяемые из текста программой структуризации, подвергаются анализу на предмет наличия в них чужой речи; решение отнести тот или иной абзац к одному из типов письменной речи принимается по завершении анализа всех предложений данного абзаца.

Построение текста с чужой речью может осуществляться двумя способами:

1) чужая речь оформляется в виде отдельных реплик участников разговора; в тексте такие реплики начинаются с новой строки;

2) чужая речь входит в состав абзаца авторского повествования; в этих случаях чужая речь обычно заключается в кавычки.

Построение алгоритма расслоения для первого варианта не вызывает особых затруднений: ситуация <АО+тире> в начале строки и отсутствие двоеточия в конце предыдущей строки являются необходимым и достаточным условием фиксации начала абзаца прямой речи. То же, безусловно, справедливо и для случая <АО+кавычки>.

Второй вариант использует кавычки и, если бы те же кавычки не служили средством оформления заимствований (названий, цитат, пословиц, поговорок и др.), то и здесь алгоритмы расслоения был бы тривиален. Ср.:

Петушков встрепенулся,— а солдат вытянулся, пожелал ему «здравья» и вручил ему большой конверт, запечатанный казенной печатью.

(И. С. Тургенев. Петушков)

Это предложение, конечно же, не следует относить к предложениям с прямой речью, следовательно, алгоритмически мы должны отличать подобные случаи и относить их к авторской речи.

Решить эту проблему можно лишь одним способом: проанализировать и формализовать правила графического оформления прямой речи с использованием кавычек. Другой вариант решения, кажущийся, на первый взгляд, более перспективным, — уточнение и формализация отличительных признаков оформления предложений с заимствованиями — алгоритмизации не поддается. Причина здесь одна — отсутствие каких-либо ограничений на использование в тексте цитат и прочих заимствований, что в отношении поиска отличительных пунктуационных особенностей делает задачу бесперспективной.

Для описания правил выявления прямой речи, заключенной в кавычки, воспользуемся ранее введенными формализмами и некоторыми новыми: через S будем обозначать некоторую пунктуационно-знаковую ситуацию в середине предложения, знаком $\vee$ обозначим логическую операцию «или», а специальным словом nil — отсутствие кавычек в N- или F-ситуациях. Тогда правила фиксации прямой речи, оформленной с помощью кавычек, могут быть сведены к четырем случаям, представленным в табл. 2.

Вариант 1 реализует тот случай, когда прямая речь разрывается словами автора на две части. Здесь S-ситуация должна содержать два тире, разделенных некоторым текстом:

«Может быть,— подумал я,— ты оттого именно меня и любила: радости забываются, а печали никогда...»

(М. Ю. Лермонтов. Герой нашего времени)

Таблица 2. **Правила выявления прямой речи (при использовании кавычек)**

F-ситуация <конец предложения>	<Z» >V<«Z>		nil	
N-ситуация <начало предложения>	<«E>	nil	<«E>	
S-ситуация <середина предложения>	<—...—>	<:«E>	<», —>	<Z», —>
Вариант	1	2	3	4

Вариант 2 реализует тот случай, когда прямая речь находится в конце предложения и следует за словами автора:

Староста сперва проворно соскочил с лошади, поклонился барину в пояс, промолвил: «Здравствуйте, батюшка Аркадий Павлович».

(И. С. Тургенев. Бурмистр)

Варианты 3 и 4 реализуют случаи, когда прямая речь находится в начале предложения, а слова автора следуют за ней. Ср. вариант 4:

«Скорей, скорей в город за лекарем!» — кричал Владимир.

(А. С. Пушкин. Дубровский)

Те же правила выявления прямой речи применимы и в более сложных случаях, а именно тогда, когда прямая речь состоит не из одного, а из нескольких предложений. При определении типа абзаца те предложения, которые расположены внутри кавычек, рассматриваются как одно целое, а знаки препинания «.», «?» и «!» в расчет не принимаются. Ср. вариант 1 и 3:

«Может быть, мы никогда больше не увидимся, — сказал он мне. — Перед разлукой я хотел бы с вами объясниться».

(А. С. Пушкин. Выстрел)

«Наплюй на него, бабонька. Была бы шея, а ярмо будет», — с нескрываемым сожалением посоветовала одна.

(М. А. Шолохов. Тихий Дон)

Разработанные и представленные выше формальные правила выявления прямой речи охватывают три наиболее употребительных варианта построения предложений с вводящими словами автора, а именно, когда слова автора: 1) предшествуют прямой речи, 2) следуют за прямой речью, 3) включаются в прямую речь, разделяя ее на части. Существует и еще один, более редкий вариант, когда вводящие слова автора включают в себя прямую речь. Однако и этот вариант формализуем путем некоторого усложнения решающих правил:

На вопрос мой: «Жив ли старый смотритель?» никто не мог дать мне удовлетворительного ответа.

(А. С. Пушкин. Станционный смотритель)

И только когда он шептал: «Мама! Мама!» — ему становилось как будто легче...

(А. П. Чехов. Степь)

Этих примеров вполне достаточно, чтобы убедиться в следующем: рассматриваемые нами конструкции содержат вводящие слова автора, которые находятся левее прямой речи и отделяются от нее двоеточием. Поэтому поисковая процедура для предложений с вводящими словами автора (вариант 5) выглядит следующим образом:

$$(S1 = <: «E>) \ \& \ ((S2 = <«, >) \vee (S3 = <Z»>)),$$

где знаком & обозначена логическая операция «и», а ситуация S1 должна предшествовать появлению в тексте ситуации S2 или ситуации S3. Поэтому для последнего варианта правило формализованного выявления прямой речи имеет вид

$$(F = nil) \ \& \ (N = nil) \ \& \ (S1) \ \& \ (S2 \vee S3).$$

Итак, мы сформулировали и формализовали правила, позволяющие строить алгоритмы расслоения текстов художественной прозы по типам письменной речи. Еще раз повторим, что основными единицами текста здесь являются абзац и предложение: каждое предложение относится к тому варианту письменной речи, к которому принадлежит и весь абзац.

Основными вариантами письменной речи, вычленяемыми процедурой расслоения, являются: авторская речь, прямая речь, «смешанная» речь и «вложенная» прямая речь. Последние три являются детализацией чужой речи:

а) прямая речь — это речь собственно диалогическая, без каких-либо вкраплений авторской речи;

б) «смешанная» речь включает в себя прямую речь внутри абзаца авторской речи;

в) «вложенная» прямая речь содержит прямую речь внутри абзаца прямой речи. Пример «вложенной» речи:

— Не радуйся, однако. Я как-то вступил с ней в разговор у колодца случайно; третье слово ее было: «Кто этот господин, у которого...»

(М. Ю. Лермонтов. Княжна Мери)

Все три варианта чужой речи характеризуют, на наш взгляд, особые стилистические приемы, которыми в той или иной степени владеет и которые использует в своем творчестве каждый автор, приемы, которые должны изучаться как в отдельности, так и в своем единстве, интегрируясь в индивидуально-авторский и жанрово-стилистический набор методов, приемов, средств письма.

Таблица 3. **Стилистические характеристики текстов**

Характеристика	Кол-во абзацев	Кол-во предложений
	текст 1/2	текст 1/2
Авторская речь . .	21/22	51/31
Чужая речь	18/42	117/117
прямая	9/34	39/92
смешанная . .	6/7	34/22
вложенная . .	3/1	44/3
Итого:	39/64	168/158

Небольшой фрагмент сопоставительного анализа двух рассказов (табл. 3) демонстрирует некоторые стилистические особенности авторской манеры письма А. С. Серафимовича (рассказ «Политком» — текст 1) и С. П. Подьячева (рассказ «Понял» — текст 2). Структуризация и расслоение текстов выполнены в автоматическом режиме системой DISSKOTE. Объем этих двух рассказов практически одинаков (по 7 страниц текста), другие же характеристики могут, без сомнения, послужить отправной точкой для глубоких стилистических изысканий.

ЗАКЛЮЧЕНИЕ

Изучение взаимовлияния разных типов речи в рамках единого целого, использование в этих целях различных концепций, методов и алгоритмов составляют одну из центральных и актуальнейших проблем лингвистики текста. Задача компьютерной лингвистики в этом неисчерпаемом творческом поиске — вооружить исследователя всем необходимым ему материалом, подготовить счетные данные и предложить широкий набор вычислительных процедур, способных в своей совокупности вести проверку гипотез, подтверждая либо опровергая все более тонкие и глубокие филологические наблюдения. Такое видение целей и задач лежит в основе разработки и создания компьютерной системы ЛИНДА — экспертной филологической системы, многофункциональной по своему назначению и «дружественной» по отношению к своим пользователям.

Итак, мы завершили обсуждение методов и конкретных вариантов формирования однородных счетных единиц текста. Этот, казалось бы, вспомогательный, подготовительный этап на самом деле должен планироваться и осуществляться с особой тщательностью и научной обоснованностью, ибо никакие, даже самые тонкие, лингвостилистические процедуры не гарантируют точность и надежность результатов при обработке некорректно сформированных (заданных, выбранных) исходных

счетных единиц текста. Несомненно также и то, что метод структуризации и расслоения текстов, разработанный для формирования однородных совокупностей текстов художественной прозы, должен, во-первых, совершенствоваться с учетом практического опыта его использования и, во-вторых, расширяться с целью своего распространения на другие жанрово-стилистические варианты письменной речи, например, поэтические тексты и драматические произведения. Это потребует разработки более тонких алгоритмов с учетом необходимости решения проблемы «дефис — перенос», анализа стандартных сокращений, авторских ремарок, различных жанрово-стилистических вкраплений.

Литература

Ван Тассел Д. Стиль, разработка, эффективность, отладка и испытание программ. М., 1981. 320 с.

Гальперин И. Р. Текст как объект лингвистического исследования. М., 1981. 140 с.

Мартыненко Г. Я. Основы стилеметрии. Л., 1988. 174 с.

Марусенко М. А. Атрибуция анонимных и псевдоанонимных литературных произведений методами теории распознавания образов. Л., 1990. 240 с.

Машинный фонд русского языка: Идеи и суждения / Под ред. Ю. Н. Караулова. М., 1986. 240 с.

Попов Э. В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. М., 1987. 288 с.

Тулдава Ю. А. Проблемы и методы квантитативно-системного исследования лексики. Таллинн, 1987. 204 с.

Шапиро А. Б. Современный русский язык: Пунктуация. М., 1974. 287 с.

АТРИБУЦИЯ АНОНИМНЫХ И ПСЕВДОНИМНЫХ ТЕКСТОВ МЕТОДАМИ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

При решении сложных научных проблем полезным и испытанным приемом является объединение разнотипных на первый взгляд задач в более широкие классы, позволяющее искать методы их решения на новой методической базе, на более высоких уровнях абстракции, используя достижения других, порою отнюдь не родственных наук. Так, при более внимательном науковедческом анализе таких традиционных филологических задач, как атрибуция анонимного или псевдонимного текста, определение или уточнение датировки того или иного текста, типология жанров и функциональных стилей, становится очевидно, что, по существу, во всех вышеперечисленных случаях мы имеем дело с одной и той же научной задачей — с объединением объектов в однородные классы, либо с отнесением объекта (группы объектов) к какому-либо классу, либо с определением меры близости/удаленности между разными объектами (группами объектов). Представляется целесообразным на основе общности этих задач, объединив их под уже введенным в научный обиход понятием «распознавание образов», разрабатывать и единые методы их решения.

При всей актуальности этих проблем и несмотря на то, что работа над ними ведется со времен филологов-классиков, нынешнее состояние вопроса таково, что не позволяет однозначно и с фиксированной достоверностью производить атрибуцию анонимного или псевдонимного текста или другие классификационные процедуры. Так, критическая статья о пьесе А. Н. Островского «Бедность — не порок» помещена в собраниях Н. Г. Чернышевского и Н. А. Добролюбова. В книге В. О. Ключевского «Неопубликованные произведения», выпущенной в 1983 г., помещено стихотворение «Вы, философ, дайте руки. . .», на самом деле принадлежащее перу Аполлона Майкова и полностью опубликованное в майковском томе «Библиотеки поэта» под на-

званием «Тарантелла». В критике оспаривается желание некоторых литературоведов приписать В. Г. Белинскому большое число неподписанных критических статей в тех журналах, где он сотрудничал. Аналогично обстоит дело и в отношении публицистических статей М. Е. Салтыкова-Щедрина, и в отношении многих материалов «Колокола», приписываемых А. И. Герцену. Даже в академических собраниях сочинений появляются иногда разделы, содержащие произведения, принадлежность которых данному автору лишь предполагается. Особую важность эта проблема имеет для изучения творчества прогрессивных деятелей русской литературы XIX в., по цензурным соображениям вынужденных публиковать свои работы анонимно или под псевдонимом.

Произведения анонимного или спорного авторства существуют в любой литературе. До настоящего времени процедура атрибуции таких текстов сводилась к тому, что филолог-литературовед, хорошо знакомый с периодом, в который был написан спорный текст, и с творчеством писателей, работавших в то время, пытается найти те или иные доказательства, позволяющие приписать данный текст известному ему автору. При этом самым важным являлись свидетельства внешнего характера, такие, как биографические данные предполагаемого автора, его пребывание в данном месте, круг его знакомств, переписка, политические взгляды и т. п. Пользуясь этим методом, разные исследователи зачастую приходят к совершенно противоположным выводам относительно предполагаемого автора анализируемого текста. Необходимо учитывать и субъективный фактор, который оказывает негативное воздействие в двух планах: с одной стороны, исследователи выбирают в качестве оснований для атрибуции те критерии, которые лучше «укладываются» в их гипотезу, с другой — проявляется отчетливая тенденция приписывать спорные тексты перу наиболее крупных и значительных писателей. Эта тенденция, о которой с тревогой писал еще В. Я. Брюсов в статье «Новооткрываемый Пушкин», вызывала тревогу и у В. В. Виноградова, разделявшего его беспокойство по поводу этого «болезненного пристрастия русских литературоведов».

Крупнейший русский лингвист академик В. В. Виноградов (Виноградов, 1960) дал следующую типологию существующих в литературоведении принципов атрибуции.

I. Субъективные: а) субъективно-коммерческие; б) субъективно-конъюнктурные; в) субъективно-эстетические; г) субъективно-психологические; д) субъективно-идеологические.

II. Объективные: а) документально-рукописные (археологические); б) исторические (биографии, сведения современников и пр.); в) историко-идеологические и сопоставительно-идеологические; г) историко-стилистические; д) художественно-стилистические; е) лингвостилистические.

Используемые по отдельности, эти принципы не позволили приблизиться к решению проблемы атрибуции в целом. Невысокая эффективность большинства подходов к установлению авторства текстов и реальное положение дел с атрибуцией в нашем литературоведении, которое трудно считать нормальным, вызвали пессимистическое отношение к принципиальной возможности решения этой проблемы.

Новые перспективы в этой области открываются при комплексном подходе к проблеме атрибуции, когда для признания принадлежности анализируемого произведения данному автору необходимо учитывать доводы по меньшей мере трех основных групп: а) биографические; б) идеологические; в) стилистические. Чем большее число разных методов используется в процессе атрибуции, тем больше его точность и надежность, достоверность окончательных выводов. Учет доказательств каждой группы является необходимым, но отнюдь не достаточным для достоверной атрибуции текста. Даже документальные данные, являющиеся, по общему мнению, наиболее авторитетным и достоверным источником, нередко бывают весьма противоречивы. Так, в упомянутом выше случае неправильной атрибуции В. О. Ключевскому стихотворения А. Майкова источником ошибки послужил найденный в бумагах В. О. Ключевского автограф этого стихотворения, оказавшийся впоследствии рукописным списком понравившегося ему чужого стихотворения. Использование интуиции исследователя, несмотря на неточности, субъективизм и элементы произвола, имеет большие преимущества по сравнению с формальными методами анализа, заключающиеся в использовании широкого социального контекста, «тезауруса культуры».

Большие надежды возлагались на использование в целях атрибуции математико-статистических методов, пионером которых был Н. А. Морозов, напечатавший в 1915 г. работу «Лингвистические спектры» с подзаголовком «Средство для отличия плагиатов от истинных произведений того или другого известного автора. Стилеметрический этюд». В отличие от предшественников, филологов-классиков, опиравшихся при атрибуции на частоту употребления знаменательных слов, Н. А. Морозов считал, что гораздо более показательными для индивидуального стиля писателя являются служебные слова, не связанные с темой, предметом речи, содержанием книги. При всей ограниченности такого подхода, связанного с возможностями самого метода лингвистических спектров, работы Н. А. Морозова внесли значительный вклад в разработку объективных методов атрибуции прежде всего тем, что была показана необходимость анализа и других уровней языковой структуры помимо лексического. Отметим, что при подделке или имитации текста легче всего добиться лексического сходства, в то время как имитация латентных структурных характеристик невозможна в силу

их неэксплицитного характера. Этим обстоятельством пользуются писатели-пародисты, воспроизводящие в пародии прежде всего лексический строй пародируемого произведения.

Несмотря на определенные успехи в применении математико-статистических методов для целей атрибуции в целом, как отмечал акад. В. В. Виноградов, еще нет единодушной точки зрения на их место и роль в общей системе методов атрибуции. Более того, значительная часть филологов весьма скептически относится к необходимости и возможностям стилистической экспертизы в целях атрибуции. Вместе с тем субъективные методы определения автора уже отжили свой век. Правильные атрибуции, выполненные с применением субъективно-эстетических, субъективно-психологических и субъективно-идеологических приемов, являются результатом интуиции или личной одаренности исследователей, а не итогом правильной методики установления авторства.

Уход теории атрибуции в своеобразный методологический тупик объясняется двумя основными причинами. Во-первых, задачи атрибуции в соответствии с целью исследования традиционно решались в рамках литературоведения, т. е. традиционной филологической дисциплины, в то время как необходимый для этого научный аппарат по своему содержанию далеко выходит за пределы традиционной лингвистики, что затрудняет, а зачастую делает невозможным его использование специалистами-литературоведами.

Во-вторых, эффективное использование объективных методов атрибуции на основе анализа языка и стиля сдерживается недостаточной изученностью последних. Необходимо планомерное и систематическое накопление данных в этой области (конкретные специфические признаки стиля, слога, словоупотребления, синтаксические, фразеологические и другие особенности языка), позволяющих производить атрибуцию с применением статистико-вероятностных методов оценки достоверности. Пока же встречающиеся в работах по атрибуции языково-стилистические характеристики носят случайный, бессистемный характер. Как отмечал М. П. Штокмар (1960), нередко за специфику языка того или иного писателя выдается то, что является характерным для языка всей его эпохи. В то же время если отдельные элементы языка и стиля писателя доступны для стилизации и заимствования, то система стиля является уникальной и невоспроизводимой.

Выход из этого методологического тупика лежит на пути использования средств и методов прикладной лингвистики, а также четкого разграничения сферы действия и возможностей традиционных филологических методов, их комплексного сочетания со статистико-вероятностными методами, разработанными для решения классификационных задач и применяемых для задач атрибуции на основе межнаучного переноса.

В свете этих взглядов вся процедура атрибуции эвристически делится на два этапа: 1) этап формирования гипотезы; 2) этап проверки гипотезы и интерпретации результатов. Этап формирования гипотезы выполняется при помощи традиционных филологических методов атрибуции с учетом как субъективных, так и объективных данных. На этом этапе может быть сформулирована гипотеза о принадлежности текста M_1 писателю A. Может выдвигаться гипотеза о времени написания текста M_1 на основе его языковых и стилистических связей с текстами $M_2, M_3, \ldots, M_n$ писателя A, датировка которых известна, и т. п. Этап проверки гипотезы и интерпретации результатов выполняется методами прикладной лингвистики с использованием процедур статистико-вероятностного анализа, теории классификации, теории распознавания образов. Как будет показано ниже, выполнение этого этапа возможно только с использованием ЭВМ и требует разработки соответствующего программного обеспечения.

Данная процедура имеет итеративный характер: в случае опровержения нулевой гипотезы о принадлежности текста M_1 писателю A должна быть сформулирована новая гипотеза о принадлежности этого текста писателю B, т. е. возврат к первому этапу процедуры, и т. д. Нулевой гипотезой называется основная выдвинутая гипотеза. Обычно нулевые гипотезы утверждают, что различие между сравниваемыми величинами (параметрами или функциями распределения) отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями выборки. В случае определения датировки последовательно проверяются нулевые гипотезы о сходстве текста M_1 с текстами $M_2, M_3, \ldots, M_n$.

Таким образом, положительное решение по вопросу атрибуции текста может быть принято только в случае совпадения результатов филологического и математико-статистического анализа.

Связующим звеном между этими двумя областями служит принятое в стилистической диагностике операциональное определение основного понятия стилистики — стиля. В современной науке достаточно четко выражены три подхода к определению стиля: лингвистический, литературоведческий и вообще искусствоведческий, причем акад. В. В. Виноградов отмечал, что литературоведческое определение стиля пока еще не опирается на глубокую самостоятельную и общепризванную основу. Оно чаще всего колеблется между общим искусствоведческим и лингвистическим подходами, механически совмещая черты и признаки обоих. Стиль писателя, по его мнению, должен изучаться как единая, внутренне целостная система функционально согласованных средств словесного выражения. В современной лингвистике стиль рассматривается прежде всего как категория структурно-синтаксическая.

В теории классификации понятие объекта определяется через пучок свойств, необходимых для отображения предметной области. Именно значения или кортежи значений свойств обозначают объект, становятся системным, формализованным именем объекта. Можно отметить, что некоторые используемые в стилистике определения стиля значительным образом коррелируют с вышеуказанным определением. Так, например, А. И. Ефимов (1961) определяет стиль как исторически сложившуюся разновидность языка, отличающуюся специфическим составом, характером объединения и закономерностями употребления речевых средств. Таким образом, в терминах теории распознавания образов стиль можно определить как набор свойств (параметров), характеризующих состав, способы объединения и закономерности употребления речевых средств, образующих данную разновидность языка.

В дальнейшем целесообразно проводить разграничение между лингвистическим объектом p и математическим объектом x, который может быть описан заданием определенных значений конечного набора параметров. Для преобразования реального лингвистического объекта в математический объект необходимо определить такой набор измерений $\{M_1, M_2, \ldots, M_n\}$, что применение этого набора к объекту p порождает набор констант $\{x_1, x_2, \ldots, x_n\}$, представляющих значения признаков изучаемого объекта. Полученный набор величин $\{x_1, x_2, \ldots, x_n\}$ описывает математический объект x, являющийся отображением реального объекта p. Поэтому, прежде чем приступить к процедуре собственно распознавания, необходимо осуществить преобразование реального объекта (стиль в сформулированном выше определении) в математический объект, являющийся отображением реального объекта в информационном пространстве, т. е. в набор значений признаков изучаемого объекта.

Обратимся теперь к анализу процедуры распознавания текстов на естественном языке на основе языково-стилистических характеристик. Эффективная реализация этой процедуры требует осуществления некоторых логически связанных этапов.

1. Определение априорного набора индивидуальных стилистических параметров. В функциональной стилистике к настоящему времени уже накоплен значительный материал по параметризации стилей. Число параметров, предлагаемых разными исследователями для диагностики стилевых различий, составляет не одну сотню. Как показывает опыт, этот набор параметров может быть значительно сокращен путем унификации обозначений, так как большое число по-разному именуемых параметров представляет собой соотношение одних и тех же величин. На практике из всего этого разнообразия использовалось лишь несколько самых простых параметров, таких, как длина предложения, длина слова и т. д. Относительно полный перечень стилевых параметров в приведенной форме дан в нашей работе

(Марусенко, 1986). При определении исходного пространства параметров учитывалось, что стиль — это прежде всего категория структурно-синтаксическая и что основной единицей синтаксиса является предложение. Поэтому параметры, составляющие исходное признаковое пространство, берутся из работ авторов, исследовавших структуру и состав предложения квантитативными методами. Учитывая требования простоты определения значения параметра, нецелесообразно включать в исходное пространство параметры, определение значений которых требует промежуточных построений (деревьев зависимостей и т. п.).

Анализ исходных параметров показывает, что их можно разделить на две группы: а) первичные параметры, значения которых определяются непосредственно в предложении, б) производные параметры, представляющие собой отношение двух или более первичных. При этом если число первичных параметров естественно ограничено, то число производных параметров может быть увеличено в несколько раз чисто комбинаторным путем. На первый взгляд может показаться, что использование производных параметров, таких, как «коэффициент сложноподчиненности» (отношение числа подчиненных предложений к числу элементарных предложений), «коэффициент комплексности» (отношение числа подчиненных предложений к числу главных предложений) и других, более сложных, является полезным, так как позволяет увеличить объем информации, используемой при диагностике (например, в факторном анализе). На самом же деле, в отличие от истинных факторов, комбинирующих предварительно отобранные независимые информативные признаки, рассматриваемые производные параметры представляют собой комбинации одних и тех же (как в вышеприведенном примере, где оба параметра используют один и тот же первичный параметр «число подчиненных предложений»), к тому же взаимно коррелирующих первичных параметров. Поэтому целесообразно на первом этапе процедуры, при составлении априорного набора параметров, проводя унификацию обозначений, включать в используемый далее набор лишь те, которые отвечают определению первичного параметра. Таким образом, каждому лингвистическому объекту, подвергаемому анализу в целях стилистической диагностики, может быть поставлен в соответствие математический объект p, характеризуемый n-мерным вектором (n — число параметров).

2. Анализ связей исходных параметров. Специалисты по применению квантитативных методов давно обратили внимание на то, что некоторые параметры, используемые для стилистической диагностики, имеют тесно связанные между собой значения. Некоторые связи имеют простую логическую интерпретацию, например, число слов в предложении и число графем в предложении или число подчиненных предложений и число подчинительных союзов. Эти связи касаются разных уровней ана-

лиза и могут охватывать два, три и более параметров. Так, например, В. Фукс (1975) установил связь трех признаков — длины, т. е. числа слов или слогов в предложении, расчлененности, т. е. числа секций в предложении, и вложенности, т. е. суммы значений рангов. И. П. Севбо (1981), изучая структуры синтаксических деревьев, определила прямую зависимость между длиной пути в графе и числом слов во фразе. Отмечены зависимости между размерами предложения и абзаца и т. д. Все эти зависимости, по мнению лингвистов, включаются в систему «скрытой упорядоченности» в объеме синтаксических единиц и являются проявлениями общей тенденции к «изоквантности», выражающейся в пропорциональности между объемами синтаксических единиц разных уровней. В ряде случаев эти связи получили количественное выражение при помощи методов корреляционного анализа, однако в лингвистике еще не применялось положение о том, что мера связи параметров может не только использоваться для оценки степени отклонения их совместного распределения от независимости, но и характеризовать возможность прогнозирования значений одного параметра по значению другого.

3. Выбор информативных параметров. Если информационное пространство построено на параметрах, то в нем может быть выделено пространство меньшей размерности, состоящее из необходимого и достаточного числа параметров для отнесения объекта к классу. Следовательно, задача оптимального свертывания признакового пространства размерности n может быть сформулирована как нахождение m признаков $(m<n)$, преобразующих реальный объект p в математический объект $M_{(p)}=M_{1(p)}, \ldots, M_{m(p)}$. Желательно, чтобы $M_{(p)}$ определялся малым числом признаков и чтобы измерение этих признаков было относительно простым. Снижение размерности исходного пространства важно с трех точек зрения: во-первых, если объем лингвистического объекта сравним с размерностью признакового пространства, то среди признаков могут быть «шумовые», удаление которых только улучшит работу распознающего алгоритма; во-вторых, упрощение математической модели объекта позволяет лучше понять механизм явления и дать содержательную интерпретацию результатов; в-третьих, снижение размерности позволяет существенно снизить вычислительные трудности, связанные с машинной обработкой значительных массивов информации большой размерности. Кроме того, это дает возможность уменьшить объем данных на этапе подготовки к вводу в ЭВМ и тем самым снизить число ошибок.

Число параметров, используемых в алгоритме распознавания, зависит от конкретных условий реализации каждого эксперимента, поэтому минимальное необходимое и достаточное число признаков определяется требуемыми характеристиками точности и устанавливается экспериментально. Процедура опти-

мального свертывания признакового пространства может быть сведена к получению набора параметров, ранжированных по убыванию их индивидуальной информативности, под которой понимается способность не снижать достоверность классификации при переходе от исходного набора к набору меньшей размерности. Из этого ранжированного ряда параметров исследователь, исходя из требований точности, возможностей и условий эксперимента, может формировать математический объект, образованный n параметрами ($n \geqslant 1$), такой, что информативность этих параметров будет превышать информативность любых других n параметров, взятых из того же исходного пространства.

Различные процедуры, используемые для свертывания параметрического пространства, основаны на анализе корреляционной матрицы связи параметров, элементами которой являются парные коэффициенты корреляции Пирсона:

$$r = \frac{N \sum x_{i1} x_{i2} - \left(\sum x_{i1}\right)\left(\sum x_{i2}\right)}{\left[N \sum x_{i1}^2 - \left(\sum x_{i1}\right)^2\right]^{1/2} \left[N \sum x_{i2}^2 - \left(\sum x_{i2}\right)^2\right]^{1/2}},$$

где N — число пар значений; x_{i1}, x_{i2} — значения парных признаков.

В настоящее время известно несколько конкурирующих методов отбора информативных параметров на основе корреляционной матрицы. Одним из наиболее простых, допускающих ручную реализацию, является метод корреляционных плеяд (Марусенко, 1990). В основе этого метода лежит предположение, что в основе сильной связи между параметрами лежит некоторый непосредственно ненаблюдаемый внутренний фактор. Поэтому параметры группируются в корреляционные плеяды, каждая из которых объединяет близкие между собой параметры, а в разные плеяды попадают параметры, слабо связанные между собой. После разбиения параметров на плеяды можно оценить, сколько «представителей» от каждой группы нужно отобрать для последующего анализа. Однако при большом числе параметров количество связей настолько велико, что для выделения корреляционных плеяд приходится прибегать к формальным методам расщепления признакового пространства.

Более алгоритмизируемым по сравнению с предыдущим является метод, основанный на критерии минимизации потерь информации, поскольку все необходимые данные могут быть получены из самой совокупности полученных связей, т. е. минимизации функционалов Σr_i, $\Sigma |r_i|$ или Σr_i^2.

Еще более мощным методическим приемом является метод определения потерь информации при последовательнном исключении каждого параметра из совокупности. При этом определяется общее количество шенноновской информации относительно связей всех параметров, представленных в корреляцион-

ной матрице, а затем определяется величина информации, остающаяся в корреляционной матрице при вычеркивании i-го столбца и i-й строки.

Шенноновская мера индивидуальной информативности параметров определяется соотношением

$$I_i = 0{,}5 \ln \frac{\det A_i \det B_i}{\det C},$$

где $\det C$ — определитель исходной матрицы, $\det A_i$ — определитель остаточной матрицы, полученной вычеркиванием i-го параметра, $\det B$ — определитель диагонального вектора.

Зная количество информации в исходной матрице (I_0) и информацию в остаточной матрице (I_i), можно определить потерю информации при исключении i-го параметра:

$$\frac{I_0 - I_i}{I_0} \cdot 100\%.$$

Содержательным критерием существенности информативного набора параметров служит слабая корреляция отобранных параметров между собой (внутригрупповая корреляция) и сильная их корреляция с остальными параметрами, не вошедшими в этот набор (внегрупповая корреляция).

По корреляционной матрице связей параметров можно определить средние величины коэффициентов корреляции параметров, входящих в группу из m параметров, с параметрами своей группы (средняя внутригрупповая корреляция) и с остальными (средняя внегрупповая корреляция). Для определения этих величин целесообразно использовать модули коэффициентов корреляции, позволяющие лучше выражать тесноту связи при суммировании. Для этого необходимо сначала найти сумму модулей коэффициентов корреляции каждого параметра с остальными параметрами совокупности $\Sigma_{i=1}^{n} |r_{ij}|$, затем сумму модулей коэффициентов корреляции каждого параметра с параметрами своей группы $\Sigma_{i=1}^{m} |r_{ij}|$. Диагональные элементы матриц, естественно, при расчетах не учитываются. Средняя внутригрупповая корреляция каждого параметра находится по формуле

$$\bar{r}^m = \left(\sum_{i=1}^{m} |r_{ij}| - 1\right) / (m - 1).$$

Средняя внегрупповая корреляция каждого параметра с остальными находится по формуле

$$\bar{r}^{n-m} = \left[\left(\sum_{i=1}^{n} |r_{ij}| - 1\right) - \left(\sum_{i=1}^{m} |r_{ij}| - 1\right)\right] (n - m - 1).$$

На основе полученных величин можно ввести критерий «эффективности параметра»:

$$E_i = \bar{r}_i^m / \bar{r}_i^{n-m}.$$

Таким образом, если $E_i < 1$, целесообразно включать i-й па-

раметр в группу параметров, используемых в процедуре распознавания, если $E_i \geqslant 1$ — нет.

4. Визуализация данных, корректировка гипотезы. Сокращение размерности параметрического пространства и, следовательно, уменьшение объема данных, подлежащих обработке, позволяет дать содержательную лингвистическую интерпретацию обнаруженных различий между объектами и, в ряде случаев, отвергнуть или подвергнуть корректировке сформулированную ранее филологическую гипотезу. Однако для того чтобы названные операции стали возможными, необходима реализация еще одного этапа, называемого визуализацией данных. Существо процедуры может быть пояснено на примере.

Пусть имеются три лингвистических объекта, характеризующиеся пятью параметрами: П01 — число слов в цельном предложении (ЦП); П02 — число графем в ЦП; П17 — число знаменательных слов в ЦП; П32 — число подлежащих в ЦП; П42 — число согласованных определений в ЦП. Таким образом, каждый объект может быть представлен в виде пятимерного вектора:

$$M_1 = \{16{,}8;\ 111{,}833;\ 13{,}000;\ 2{,}633\},$$
$$M_2 = \{17{,}0;\ 108{,}684;\ 12{,}974;\ 1{,}737;\ 2{,}684\},$$
$$M_3 = \{15{,}387;\ 89{,}694;\ 10{,}903;\ 1{,}387;\ 2{,}306\}.$$

Представим полученную информацию в виде таблицы (табл. 1).

Таблица 1

Объект	M_1	M_2	M_3
П01	16,800	17,0	15,387
П02	111,833	108,684	89,694
П17	13,100	12,974	10,903
П32	1,400	1,737	1,387
П42	2,633	2,684	2,306

Таблица 2

Объект	M_1	M_2	M_3
П01	0,615	0,815	—0,798
П02	9,902	6,753	—12,237
П17	0,908	0,782	—1,289
П32	0,108	0,229	-0,121
П42	0,102	0,153	—0,225

Анализ этой таблицы позволяет сделать некоторые предположения о сходстве и различиях диагностируемых объектов, однако вариабельность параметров и их разнопорядковость не дают возможности сформулировать классификационную гипотезу. Определив групповые средние для каждого параметра, выразим имеющуюся информацию в отклонениях от средних (табл. 2).

Данные табл. 2 дают возможность предположить наличие в пространстве объектов разделяющей поверхности, проходящей через средние значения и делящей объекты на два класса: первый, включающий в себя объекты M_1 и M_2, находящиеся по одну сторону от поверхности, проходящей через средние, и вто-

рой, включающий в себя объект M_3, находящийся по другую сторону от этой поверхности. Однако информации, имеющейся в табл. 2, недостаточно для проверки гипотезы о сходстве или различии объектов M_1 и M_2, так как между ними нельзя провести разделяющей поверхности, как в предыдущем случае, потому что значение параметров обоих объектов отклоняется то в одну, то в другую сторону. Кроме того, данные имеют различные порядки и абсолютные величины, что исключает возможность их непосредственного сравнения.

Третьим, заключительным моментом визуализации данных является нормирование, т. е. представление значений параметров по отношению к среднеквадратичному отклонению (*табл. 3*).

Таблица 3

Объект	M_1	M_2	M_3
П01	0,051	0,067	—0,066
П02	0,126	0,086	—0,155
П17	0,096	0,083	—0,137
П32	0,093	0,197	—0,104
П42	0,043	0,064	—0,094

5. Выбор решающего правила. Решающее правило — алгоритм, обеспечивающий разделение в некотором смысле (как правило, в смысле минимума ошибок распознавания) наилучшим образом пространства признаков на области, соответствующие классы. В настоящее время известно несколько десятков методов построения решающего правила классификации, более или менее чувствительных к предположениям о распределении признаков и о типе дискриминантной функции. При построении правила классификации предполагается, что каждый класс характеризуется набором параметров, имеющих какой-то эвристически определяемый смысл. В основном это характеристики эталонов, представляемых точкой, подпространством в многомерном признаковом пространстве. Для упрощения описания классов делается предположение, что объекты каждого класса компактно расположены в признаковом пространстве в ближайшей окрестности некоторого центра (эталона).

Представление лингвистического объекта (текста) в виде многомерного вектора позволяет сформулировать задачу установления автора анонимного или псевдонимного текста как задачу нахождения расстояния между многомерным вектором, репрезентирующим атрибутируемый текст M_1, и многомерным вектором, репрезентирующим текст M_2 неизвестного автора либо совокупность текстов известного автора. Функция, выбранная для измерения этого расстояния и принятия решения о сходстве или различии этих объектов, называется решающим правилом. Строго говоря, решение о сходстве (идентичности) двух текстов и, следовательно, о принадлежности их одному автору, может приниматься только в случае, когда оба многомерных вектора совпадают, т. е. расстояние равно нулю. На практике, учитывая, что значение признака для какого-либо объ-

екта — одна из реализаций соответствующей случайной величины, устанавливается некоторый порог, при сравнении с которым принимается решение о «похожести» объекта на эталон того или иного класса. Для этого используется алгоритм классификации объектов в одномерном пространстве с помощью t-критерия:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\left(\sigma_1^2/n_1 + \sigma_2^2/n_2\right)^{1/2}},$$

причем решение о принадлежности объекта к определенному классу принимается только в том случае, если плоскости, разделяющие объекты, существуют во всех измерениях параметрического пространства. В общем случае для отнесения объекта к классу необходимо выполнение условия

$$x_i \in \Omega_A,\ t^{A, x_i}_{\text{набл}, j} \leqslant t_a,\ j = \overline{1, n},$$

$$x_i \in \Omega_B,\ t^{B, x_i}_{\text{набл}, j} \leqslant t_a,\ j = \overline{1, n}.$$

Часто в классификационных задачах в качестве меры подобия используется так называемое евклидово расстояние:

$$d_{(A, B)} = \left[\sum_{i=1}^{n} (x_{ai} - x_{bi})^2\right]^{1/2}.$$

Если C_i — центр (эталон) класса Π_i, а $D(X, C_i)$ — расстояние (близость, похожесть) между объектами X и C_i, то принимается решение:

$$X \in \Pi_1, \text{ если } D(X, C_1) < D(X, C_2),$$
$$X \in \Pi_2 \text{ — в противном случае.}$$

Принятие решения по наименьшему евклидову расстоянию основано на неявном предположении, что каждый класс характеризуется вероятностным распределением, плотность которого убывает с ростом расстояния $D(X, C_i)$ от центра C_i. Построение такой модели позволяет классифицировать любой объект, отнеся его к тому классу, для которого расстояние $D(X, C_i)$ имеет минимальное значение.

Возможен также вероятностный вариант классификации, т. е. приписывание вероятности, или степени достоверности, каждому решению. Вероятность принадлежности объекта к j-му классу вычисляется следующим образом:

$$p_{ij} = \frac{1}{d_{ij}}\left(\sum_k \frac{1}{d_{jk}}\right)^{-1},$$

где d_{ij} — расстояние между объектом и i-м классом, а d_{jk} — расстояние между объектом и остальными классами классификации.

Таким образом можно решать любую задачу отнесения объектов к уже известным классам. При наличии двух классов решающее правило вероятностного алгоритма имеет вид:

$$x_i \in \Omega_A,\ P(X_i \in \Omega_A) > 0{,}5,$$

$$x_i \in \Omega_B,\ P(X_i \in \Omega_B) > 0{,}5,$$

$$P(X_i \in \Omega_A) = P(X_i \in \Omega_B) = 0{,}5 \text{ — отказ от распознавания.}$$

Литература

Берков П. Н. Об установлении авторства анонимных и псевдонимных произведений XVIII в. // Рус. лит. 1958. № 2.

Виноградов В. В. 1) Лингвистические основы научной критики текста // Вопр. языкозн. 1958. № 2—3; 2) Проблема авторства и теория стилей. М., 1961.

Заде Л. А. Размытые множества и их применение в распознавании образов и кластер-анализе // Классификация и кластер. М., 1980.

Ермоленко Г. В. Анонимные произведения и их авторы. Минск, 1988.

Ефимов А. И. Стилистика художественной речи. 2-е изд. М., 1961.

Марусенко М. А. 1) Отбор информативных параметров в задачах стилистической диагностики // Структурная и прикладная лингвистика. Вып. 3. Л., 1986; 2) Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. Л., 1990.

Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика. Киев, 1981.

Хетсо Г., Гусставсон С., Бекман Б., Гил С. Кто написал «Тихий Дон»? М., 1989.

Фукс В. По всем правилам искусства: Точные методы в исследованиях литературы, музыки и изобразительного искусства // Моль А., Фукс В., Касслер М. Искусство и ЭВМ. М., 1975.

Штокмар М. П. Анализ языка и стиля как средство атрибуции // Вопросы текстологии. Вып. 3. М., 1960.

Vašak P. Metody určovaní autorství. Praha, 1980.

СТРУКТУРНОЕ МАТЕМАТИЧЕСКОЕ И ПРИКЛАДНОЕ СТИХОВЕДЕНИЕ
(метрика и ритмика)

Введение

Стиховедение является дисциплиной, лежащей на стыке лингвистики и литературоведения. Стихотворный текст интересует лингвистику как особого рода речь. Специфическими чертами этой речи являются, прежде всего, метрика, ритмика, рифма и строфика.

Предметом изучения структурного стиховедения являются: а) принципы организации стихотворного текста, отличающие его от прозаического; б) отражение этих принципов в тексте; в) вытекающие из них другие стороны звуковой организации текста.

Основным признаком, отличающим стихотворный текст от прозаического, является его расчлененность на отрезки, соизмеряемые в процессе их возникновения. Текст, обладающий привычными признаками стиха, например, строгим порядком в расположении ударений, но не расчлененный на такие отрезки, воспринимается как проза:

Я был готов любить весь мир, меня никто не понял: и я выучился ненавидеть.

(М. Лермонтов)

Ср. то же:

Я был готов любить весь мир, —
Меня никто не понял: и
Я выучился ненавидеть.

Текст, не обладающий четко выраженными признаками упорядоченности, но расчлененный на соизмеряемые по ходу возникновения отрезки, воспринимается как стихи:

Три раза я его видел лицом к лицу.
В первый раз шел я по саду,
посланный за обедом товарищами,

и, чтобы сократить дорогу,
путь мимо окон дворцового крыла избрал я...

(М. Кузмин)

Соизмеряемые отрезки, на которые делится стихотворный текст, называются стихами или стихотворными строками.

В типичном случае разделение на стихи и их соизмерение осуществляются на основе метра. Метр есть принцип организации слоговой структуры стихотворных строк, достаточно выраженный для того, чтобы создавать: а) ожидание подтверждения в следующих стихах; б) специфическое переживание перебоя при его нарушении.

И в суму его пустую
Суют грамоту другую.

(А. Пушкин)

Во второй строке намеренное нарушение метра. Художественный прием рассчитан на то, что читатель отреагирует на это нарушение.

В русской поэзии имеется пять классических метров: ямб, хорей, дактиль, амфибрахий и анапест. В их основе лежит периодическое чередование сильных (СМ, —) и слабых (сл, ⏑) мест. Период, лежащий в основе метра, называется стопой. В ямбе это группа слСМ (⏑—), в хорее — СМсл (—⏑), в дактиле — СМслсл (—⏑⏑), в амфибрахии — слСМсл (⏑—⏑), в анапесте — слслСМ (⏑⏑—).

В тексте различие между СМ и сл проявляется количественно и качественно. Количественное различие выражается в том, что на СМ ударения попадают часто, а на сл — редко. Качественное различие проявляется в так называемом запрете на переакцентуацию: на сл не должно находиться ударение слова, падающее безударным слогом на СМ. В тексте такие случаи редки. Их появление вызывает ощущение перебоя:

Литейщик приложил щеку,
Целясь, к морозному прикладу...

(Э. Багрицкий)

Те слоги стихотворной строки, которые падают на СМ, называются сильными, а те, которые падают на сл, — слабыми.

Схема расположения сильных и слабых слогов в стихотворной строке называется размером. Размер определяет частную разновидность метра: двух-, трех-, четырех-, пяти- и т. п.-стопный ямб, хорей, дактиль, амфибрахий, анапест. Например:

Что за скука, что за горе наше бедное житье!
День приходит, день проходит — видно, слышно все одно...

(А. Пушкин)

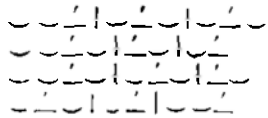
Эти стихи написаны восьмистопным хореем. Если разделить каждую строку на две, получим четырехстопный хорей:

Что за скука, что за горе
Наше бедное житье!
День приходит, день проходит —
Видно, слышно все одно...

Кроме классических метров в русской поэзии существуют и другие. Они характеризуются иным принципом организации СМ и сл. Дольник, например, построен так, что число сл между двумя СМ колеблется от одного до двух:

Небеса унылы и низки,
Но я знаю — дух мой высок.
Мы с тобою так странно близки,
И каждый из нас одинок...

(З. Гиппиус)



Акцентное противопоставление слогов является не единственным конструктивным признаком, который может быть положен в основу метра. В поэзии других народов в основу метра могут быть положены конструктивные признаки, не свойственные русскому стиху. Так, один из главных метров античной классической поэзии — гекзаметр — в качестве конструктивного признака использовал долготное противопоставление слогов. Это шестистопный дактиль, однако СМ выделяются здесь не ударностью, а долготой. Они могли заполняться только долгими слогами, в то время как два соседние с ним сл заполнялись либо двумя краткими, либо одним долгим слогом. Так как в русском языке долгота не является фонологическим признаком, то при переводе долготное противопоставление слогов заменялось акцентным. На СМ русского гекзаметра всегда стоит ударение, а слабый промежуток заполняетя одним или двумя безударными слогами:

Ласточка, ласточка, как я люблю твои вешние песни!
Милый твой вид я люблю, как весна и живой и веселый!
Пой весны провозвестница, пой и кружись надо мною;
Может быть, сладкие песни и мне напоешь ты на душу.

(Н. Гнедич)

Наряду с метром имеются неметрические принципы членения и соизмерения стихов. В действительности существуют разные стадии отклонения от метризованного текста. Наименее организованным является так называемый свободный стих — стих, в котором нет рифмы и обоих указанных ранее признаков метра (см. выше стихи М. Кузмина).

В зависимости от того, какой конструктивный признак положен в основу соизмерения стихов и как он их организует, различаются системы стихосложения. В русской поэзии выделя-

ются три системы стихосложения: силлабическая, силлабо-тоническая и тоническая.

В первой системе в основу соизмерения положен слог. Строки соизмеряются по числу входящих в них слогов. В типичном случае все строки стихотворения имеют одинаковое количество слогов:

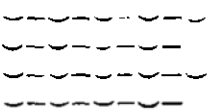
Никаков плод, все кричат, не видим науки;
Ученых хоть голова полна — пусты руки...

(А. Кантемир)

Во второй системе в основу соизмерения наряду со слогом положено акцентное противопоставление слогов. Стихи организованы так, что сильные слоги не меняют своих *позиций* в последовательности строк. Количество слабых слогов между соседними сильными и перед первым из них остается постоянным во всех строках. В типичном случае стихи подчиняются тому или иному размеру классического метра:

А сам закат в волнах эфира
Такой, что мне не разобрать,
Конец ли дня, конец ли мира,
Иль тайна тайн во мне опять.

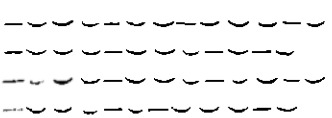
(А. Ахматова)



Силлабо-тоника отделяется от тоники переходными формами. К ним относятся дольник и тактовик. По своей организации эти стихи отличаются от силлабо-тонических тем, что позиции сильных слогов в строке становятся переменными. Число слабых слогов между двумя сильными может меняться от одного до двух в дольнике (см. выше стихи З. Гиппиус) и от одного до трех в тактовике.

Я —
безоговорочно и бесповоротно —
капля
в океане моего народа.
Истовом,
березовом,
бурлацком,
бунтарском!
В стонах и частушках.
Счастье и мытарстве...

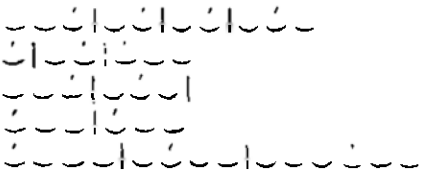
(Р. Рождественский)



В тонической системе в основу соизмерения положено слово. Внутренняя организация стиха более аморфна, чем в переходных формах. Число безударных слогов между двумя ударными может колебаться столь свободно, что понятия СМ и сл в этой системе исчезают. Метра нет. Стихи различаются обычно по количеству ударений — двух-, трех-, четырех- и т. д. ударные:

Обдают дождем дела бумажные,
чуть войдешь в здание:
отобрав с полсотни —
самые важные! —
служащие расходятся на заседания.

(В. Маяковский)



Раздел стиховедения, занимающийся изучением метра и других принципов, на которых базируется членение текста на стихи и их соизмерение, называется метрикой. Метрикой называется также совокупность тех свойств стихотворного текста, которые изучаются этим разделом.

Одним из главных теоретических вопросов метрики является вопрос об отношении системы стихосложения и языка: свойствен ли данному языку конструктивный признак, положенный в ее основу? Является ли принятая организация стиха по этому признаку действенной для носителя данного языка, т. е. происходит ли при восприятии соизмерение строк по этому признаку?

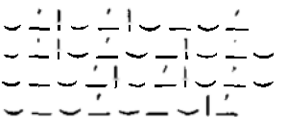
Существует, например, мнение, что силлабическая система не свойственна русскому языку, так как слог не является в нем достаточно отчетливым. Эта система была перенесена в русскую поэзию из польского стихосложения Симеоном Полоцким. Некоторые исследователи считают, что она продержалась долго в русском стихе (с 1660 примерно до 1740 г.) потому, что стихи читались с искусственным выделением слогов. Это мнение встречает возражения со стороны других ученых.

Изучаемые метрикой принципы организации обычно отображаются в тексте в виде слоговых конструкций. Эти конструкции определяют ритмическую структуру (РС) стихотворного текста. Так, в русской поэзии РС стихотворного текста будет соответствующая ему последовательность ударных и безударных слогов, разделенная словоразделами:

Она еще не родилась,
Она — и музыка, и слово

И потому всего живого
Ненарушаемая связь...

(О. Мандельштам)



Аналогичным образом определяется РС отдельного стиха (ритмическая строка), его части, последовательности стихов и т. п.

Последовательность ощущений, возникающая как синхронное отражение возобновляемого и прерываемого в каждой ритмической строке процесса ее взаимодействия с метром и с коррелятом предшествующих ритмических строк, называется ритмом:

Вагоны шли привычной линией,
Подрагивали и скрипели;
Молчали желтые и синие;
В зеленых плакали и пели.

(А. Блок)

При отсутствии метра это понятие в определении ритма опускается.

Сосуществование ритма с тем процессом взаимодействия, который он отражает, представляет собой ритменный текст.

Изучение ритменного текста относится к кругу проблем, связанных с отображением в тексте принципов его организации.

Раздел стиховедения, занимающийся изучением ритма и ритменного текста, называется ритмикой. Ритмикой называется также совокупность тех свойств стихотворных текстов, которые изучаются этим разделом.

Одним из главных теоретических вопросов ритмики является вопрос об отношении ритма и языка: определяется ли характер ритменного текста непосредственным воздействием языка или вербальное наполнение преимущественно подстраивается под ритм? Уместно вспомнить в связи с этим широко известное высказывание Маяковского:

«...Сначала стих „Есенину" просто мычался, приблизительно так:

та-ра-ра / ра-ра / ра, ра-ра, ра / ра-ра /
ра-ра-ра / ра-ра-ра / ра-ра-ра-ра-ра-ра
ра-ра-ра / ра-ра-ра-ра-ра-ра-ра /
ра-ра-ра / ра-ра-ра /ра-ра / ра / ра-ра

Потом выясняются слова».

В области метрики и ритмики можно выделить четыре основных направления структурного стиховедения, различающихся подходом к изучению предмета: 1) дескриптивное; 2) аксиоматическое; 3) генеративное; 4) реконструктивного моделирования.

Структурное стиховедение широко опирается на математические методы исследования. В основном они связаны со статистической обработкой текстовых данных. Используемый аппарат может переноситься в стиховедение в готовом виде или разрабатываться специально применительно к его задачам. Го направление исследований, которое занимается разработкой и теорией математических методов для решения задач стиховедения, называется математическим стиховедением. Направление стиховедческих исследований, обслуживающее внешние по отношению к стиховедению задачи, называется прикладным стиховедением. Математическое и прикладное стиховедение могут быть связаны с любым из перечисленных выше четырех направлений.

Рассмотрим теперь более подробно сущность выделенных направлений структурного стиховедения, основные задачи и некоторые результаты исследований.

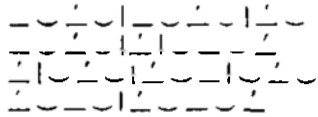
ДЕСКРИПТИВНОЕ СТИХОВЕДЕНИЕ

Метод. Дескриптивное стиховедение как направление возникло на материале русской поэзии. Его становление началось с исследований А. Белого в области четырехстопного ямба. В дальнейшем это направление получило значительное развитие в работах Г. А. Шенгели, Б. В. Томашевского, К. Ф. Тарановского, М. Л. Гаспарова, Дж. Бейли и др. Под влиянием русского стиховедения оно получило распространение и при изучении иноязычного стиха: английского — в работах М. Г. Тарлинской, болгарского — в работах Р. Т. Кунчевой, а также в других исследованиях.

Дескриптивное стиховедение занимается изучением текстовой формы явлений и отношений между ними. Объектом описания в метрике и ритмике является РС. Основной метод исследования — дистрибутивно-статистический. Подсчитываются частоты ритмообразующих элементов в различных условиях. Результаты подсчетов подвергаются обычно дополнительному анализу, на основе чего делаются выводы о свойствах ритменного текста, метра и др.

Статистическое изучение РС основано на том, что она распадается на однородные элементы с формально выраженными индивидуальными признаками. Наиболее естественными единицами этого вида являются РС стихотворных строк. Даже при наличии единого метра в них проявляются определенные индивидуальные черты. В русских поэтических текстах, написанных классическими размерами, они вызваны, в первую очередь, различиями в расположении ударений на СМ и словоразделов:

Расплясались, разгулялись бесы
По России вдоль и поперек, —
Рвет и крутит снежные завесы
Выстуженный Северовосток.
(М. Волошин)



Единицы изучения могут быть и мельче, и крупнее ритмической строки, они могут объединять различные части строк и т. д. Выбор единиц анализа и описания существенным образом зависит от метра или заменяющих его принципов организации текста.

При изучении русского стиха, написанного классическим метром, привычными единицами анализа являются ритмические формы и ритмические варианты. *Ритмическими формами являются* ритмические строки, различающиеся положением ударений на сильных слогах, ритмическими вариантами — формы, рассеченные словоразделами. Четырехстопный ямб имеет, например, восемь ритмических форм:

1 ◡ ´ — ◡ ´ — ◡ ´ — ◡ ´ — За здравье выпьем жен румяных
2 ◡ — ◡ — ◡ ´ — ◡ ´ — Соединяясь в шумный круг
3 ◡ ´ — ◡ — ◡ ´ — ◡ ´ — Задумчивость ее подруга
4 ◡ ´ — ◡ ´ — ◡ — ◡ ´ — Науки юношей питают
5 ◡ — ◡ — ◡ ´ — ◡ ´ — Как бы палеозойской эры
6 ◡ — ◡ ´ — ◡ — ◡ ´ — И замолчавшие мечты
7 ◡ ´ — ◡ — ◡ — ◡ ´ — Бушующие листопады
8 ◡ — ◡ — ◡ — ◡ — ◡ ´ — И стереомагнитофон

Каждая форма распадается на варианты:

1 ааа ◡ ´ | ◡ ´ | ◡ ´ | ◡ ´ — Слепой тоски моей не множь
1 ваа ◡ ´ ◡ | ´ | ◡ ´ | ◡ ´ — Есть целый мир в душе твоей
1 аав ◡ ´ | ◡ ´ | ◡ ´ ◡ | ´ — Еще земли печален вид

а — слово не имеет заударных слогов, в — слово имеет один заударный слог.

В качестве обобщенной первичной характеристики стиха такого рода нередко используется профиль ударности — последовательность чисел, показывающая частоты ударений на каждом СМ. Вот, например, профиль ударности «Евгения Онегина»:

СМ	I	II	III	IV
частота	0,844	0,899	0,431	1,000

Первичное описание РС в дескриптивном стиховедении не обязательно сводится лишь к статистике выделяемых единиц. Оно может предполагать и более сложную обработку. В конечном итоге выбор того или иного метода описания определяется характером изучаемой задачи

Одним из важных исследовательских средств дескриптивного стиховедения является сравнительный анализ. Он осуществляется в двух направлениях: а) сравнение с показателями дру-

гих текстов, б) сравнение с теоретическими показателями, рассчитанными на основе той или иной гипотезы о свойствах текста.

В современном стиховедении широкое распространение получило изучение ритменного текста на фоне так называемой языковой модели размера. Она рассчитывается из предположения, что поэт не накладывает никаких ограничений на ритменный текст, кроме того, что он должен укладываться в надлежащий размер. РС текста будет иметь такой же вид, какой получился бы у прозаика, если бы из всех возможных способов выражения смысла он выбирал бы тот, который не нарушает размер. Языковая модель показывает, каким будет в этом случае статистическое распределение ритмообразующих элементов. Она помогает выделить те дополнительные особенности, которые кроме метра вносит стиховая форма.

Задачи дескриптивного стиховедения в области метрики и ритмики оказываются смежными. Их разграничение основано на том, что метрика стремится к выявлению обобщающих схем, отражающих принцип организации текста, а ритмика стремится найти формы существования этих схем в текстах.

Дескриптивная метрика. Предметом изучения дескриптивной метрики являются: а) арсенал и структура метров и эквивалентных принципов организации стиха в рамках национального стихосложения; б) сравнительное изучение организационных форм в разных стихосложениях; в) исследование организационных принципов стиха в зависимости от их связи с теми или иными не входящими в сферу этих принципов факторами—языковым, жанровым и др.

Вопрос об арсенале организационных форм предполагает составление каталогов, определение пропорций и регистрацию диахронических изменений этих форм. При изучении структуры организационных форм стиха дескриптивная метрика стремится: а) найти, по возможности, схематическое выражение, обобщающее конкретные черты РС однородных текстов; б) выявить типологические виды данной организационной формы; в) проследить ее изменения в диахронии.

Рассмотрим это на примере правильного трехударного дольника русского стиха (см. Теория стиха, с. 59—106). Стих этого размера в общем виде описывается схемой: 0/2—2/1—2/1—2/0, где цифры обозначают диапазон колебаний числа слабых слогов, а черточка — сильный слог. В действительности выделяется пять типов этого размера: три основных и два переходных. Основные типы можно назвать именами поэтов, у которых они впервые отчетливо проявились: 1) есенинский (преобладают формы с промежутком в два слога между первым и вторым СМ); 2) гумилевский (преобладание форм, сочетающих двусложный и односложный интервал во внутренних слабых позициях строки); 3) цветаевский (строится на форме, сочетающей

двусложный интервал в первой внутренней позиции строки и односложный — во второй; при этом допускается и пропуск ударения на втором СМ).

1) Да богат я, богат с излишком,
Был цилиндр, а теперь его нет.
Лишь осталась одна манишка
С модной парой избитых штиблет.

2) Все проходит как тень, но время
Остается, как прежде, мстящим,
И былое, темное бремя
Продолжает жить в настоящем...

3) Завтра с западу встанет солнце!
С Иеговой порвет Давид!
Что мы делаем? — Расстаемся!
Ничего мне не говорит...

В эволюционном плане есенинский тип приблизительно соответствует *дольнику на начальной стадии его развития*, цветаевский — на конечной, гумилевский занимает промежуточное положение. Рассматривая еще и переходные типы, можно проследить последовательность преобразований дольника от есенинского и гумиловского типов к цветаевскому.

Исследование отношений между организационными принципами стиха и теми или иными внешними факторами осуществляется в связи с задачами разного рода. Языковой фактор может привлекаться, например, тогда, когда встает вопрос о причинах возникновения стиха данного типа или о диапазоне его распространения.

Одним из основных вопросов метрики русского стиха является вопрос о причинах замены силлабической системы стихосложения силлабо-тонической. Две основные точки зрения сводятся к следующему: 1) к смене систем привела постепенная тонизация силлабического стиха; 2) родоначальник русского силлабо-тонического стиха М. В. Ломоносов ввел его революционным путем под влиянием германского стихосложения.

Наиболее основательно этот вопрос изучался в исследованиях М. Л. Гаспарова. Показатели силлабических текстов в разные периоды сравнивались с соответствующей языковой моделью. На основании большого сходства между реальными и модельными показателями отдано предпочтение второй точке зрения.

Изучение метрики на основе жанровых и тематических распределений вызвано тем, что имеются исторически сложившиеся, а возможно, и органически обусловленные корреляции между этими признаками и метром. Одна из самостоятельных проблем, связанных с этим изучением, известна как проблема семантики метра. Систематические исследования в этом направлении начинаются с работы К. Ф. Тарановского о пятистопном хорее русского стиха. Им установлено, что со времени написания знаменитого стихотворения М. Ю. Лермонтова «Выхожу один я на дорогу» размер этого произведения, пятистопный хорей, тесно связан с темой пути и судьбы («Вот бреду я вдоль большой дороги» Тютчева, «...Выходила на берег Ка-

тюша» Исаковского и др.). Это объясняется как влиянием лермонтовского шедевра, так и свойствами самого размера, имитирующего, по мнению исследователя, ритм человеческого шага. Впоследствии в работах М. Л. Гаспарова и других стиховедов были обнаружены также семантические тяготения ряда других размеров.

Дескриптивная метрика получила новые данные о составе и структуре литературного и народного стиха. Она дала, с одной стороны, новое освещение известных ранее принципов организации, таких как двусложники и трехсложники русской поэзии, английский пятистопный ямб, итальянский десятисложник и др. С другой стороны, она уточнила и выявила состав и структуру различных принципов организации, не вошедших в классический канон: дольник и тактовик русской поэзии, стих В. В. Маяковского, стих «Песен западных славян» А. С. Пушкина, «Александрийских песен» М. Кузмина и др.

Дескриптивная ритмика. Дескриптивная ритмика, с одной стороны, имеет целью получить информацию о текстах. С другой стороны, она предусматривает изучение самих ритмообразующих единиц. Первое направление занимается изучением: а) универсальных и устойчивых на протяжении больших временных периодов свойств РС; б) типологических и индивидуальных свойств РС; в) ее изменениями в диахронии.

Принципиальным достижением дескриптивной ритмики в первом направлении был сам факт обнаружения устойчивых во времени и близких к универсальным явлений на материале ряда размеров. Установлена устойчивость ритмических характеристик текстов в разные периоды развития национального стихосложения. Так, в русской поэзии XVIII в. господствовал рамочный профиль ударности четырехстопного ямба — крайние сильные слоги несли ударение чаще, чем средние (рис. 1). В XIX в. господствующее положение занимает альтернирующий профиль ударности — четные сильные слоги ударяются чаще, чем нечетные (рис. 2).

Было установлено (К. Ф. Тарановский), что ритмика двухсложных размеров формируется под действием двух законов: регрессивной акцентной диссимиляции и стабилизации сильного слога после первого слабого. Первый из них формулируется так: чаще и реже ударяемые СМ чередуются через одно в направлении от последнего к первому, при этом различия между ними постепенно уменьшаются за счет ослабления ударности часто ударяемых СМ и увеличения ее на реже ударяемых. Этому закону соответствует альтернирующий профиль ударности (см. рис. 2).

Второй закон проявляется в том, что имеется тенденция к повышению частоты ударений на СМ, расположенных после первого сл (второй слог в ямбе, третий в хорее).

В определенных случаях исследователя могут интересовать

не целостные характеристики ритменного текста или его компонентов, а свойства формирующих их единиц. Это направление исследований относится к сфере лингвистики ритма. Еще

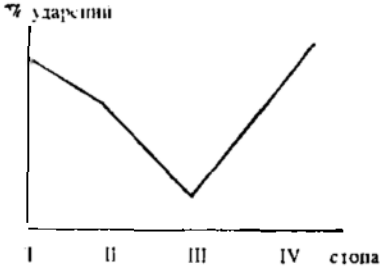


Рис. 1.

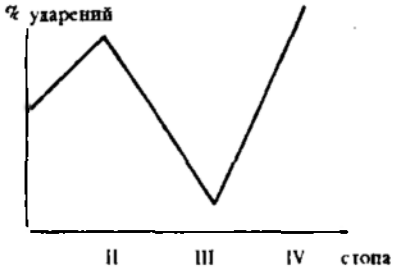


Рис. 2.

А. Белый вводил, например, темповые и другие характеристики определенных ритмических конфигураций:

«Я предлагаю следующую градацию темпов в направлении к замедлению <...>:

1) ⏑ ⏑ ⏑ — ⏑ ⏑ ⏑ — 3) ⏑ — ⏑ — ⏑ ⏑ ⏑ —
2) ⏑ ⏑ ⏑ — ⏑ — ⏑ — 4) ⏑ — ⏑ ⏑ ⏑ — ⏑ —
5) ⏑ — ⏑ ⏑ ⏑ ⏑ —».

(А. Белый, с. 266)

Ср. выше примеры форм четырехстопного ямба. Для того чтобы найти формы, соответствующие схемам А. Белого, надо обозначить каждый ударный слог знаком —, а каждый безударный, как сильный, так и слабый, знаком ⏑.

В современном стиховедении темповые, семантические и другие свойства ритмообразующих единиц изучаются на основе дистрибутивно-статистического анализа (М. А. Красноперова). При изучении их семантических предрасположенностей, например, подсчитываются частота попадания слов — носителей соответствующих РС в те или иные семантические рубрики, определяемые с помощью семантического словаря или иным способом:

В детстве, как сейчас еще помню,
Высунешься, бывало, в окно...

(Б. Пастернак)

(⏑ ⏑ ⏑ ⏑ — «движение вне»)

Реальные частоты сравниваются с теми, которые имели бы место в случае независимости семантики и РС.

Развитие современного стиховедения, в первую очередь, связано с развитием его дескриптивной отрасли. Остальные направления в той или иной мере опираются на ее результаты и методы.

АКСИОМАТИЧЕСКОЕ СТИХОВЕДЕНИЕ

Аксиоматическое стиховедение представлено работами, выполненными на материале разных стихосложений. В полном виде система аксиоматического описания метрики или ритмики включает в себя три составляющие: 1) компоненты метрики (ритмики); 2) отношения между ними; 3) правила их сочетаемости. Классическим образцом с четко выраженной методологией аксиоматического описания является работа Р. Якобсона «Аксиомы системы стихосложения (на примере мордовской песни)». Рассмотрим некоторые ее положения.

В качестве компонентов метра выделяются следующие: слог, слово, член фразы, фраза, сегмент, стих, целое. Выделенные компоненты организуются в иерархическую шкалу, в которой каждый последующий компонент содержит целое число непосредственно предшествующих: слог, слово, сегмент, член фразы, стих, целое:

Kolma pusta // s'v'etěnʒa Три пустых цветка

(7 слогов, 3 слова, 2 сегмента, 1 член фразы, 1 стих; // — граница между сегментами).

Далее рассматриваются правила оформления метрических компонентов. Организация сегмента зависит от его положения в стихе и от организации стиха. Это описывается тремя правилами. 1) Сегмент содержит не менее трех слогов и не более пяти (минимальный — три, средний — четыре, максимальный — пять). 2) Начальный сегмент не является минимальным (возможен стих 4+ . . . ; 5+ . . . ; но невозможен 3+ . . .). 3) Конечный сегмент может быть средним только тогда, когда все предшествующие сегменты стиха — средние (возможны стихи 4+4, 4+5, 4+4+4, но невозможны 3+4, 5+4 и т. п.:

Kozon' čačnês' // komul'avka? . . (4+4) Где родился хмель

Организация стиха зависит от его места в целом и от организации целого. Это описывается двумя правилами. 1) Стих содержит не менее двух и не более трех сегментов, составляющих сплошную цепь (два — короткий, три — долгий):

Ul'an s'uron', // bojaron' t' 'ejt'er' // v'id'iska (долгий)
Буду хлеб, боярышня, сеять

2) Метрически не тождественные смежные стихи являются короткими, при этом нечетный по номеру стих состоит из сегментов с одинаковым количеством слогов (т. е. 4+4, 4+5), а четный является минимальным (4+3):

Kadyk', kadyk' // c'oran'aka (4+4) Брось, брось, мой сыночек,
tabaktojak' // targamot! (4+3) Табак курить, мой сыночек

Организация целого описывается тремя правилами. 1) Целое содержит не менее двух стихов, каждый из которых метрически тождествен предпоследнему перед ним при наличии такового (смежные стихи могут быть метрически тождественны:

и не тождественны). Возможны, например, последовательности V1 V2 V3 V4, где V3 обязательно равен по метру V1, V4 — V2, а отношения между V1 и V2, V2 и V3, V3 и V4 могут быть разными: V1=V2=V3=V4 или V1#V2, V2#V3, V3#V4 (=, # обозначает равенство и неравенство по метру). 2) Если целое содержит метрически не тождественные смежные стихи, то не менее чем два из них имеют четный номер в общей последовательности, а один из подобных четных стихов завершает целое. Невозможна, например, последовательность V1 V2 V3, где V2#V3, но возможна последовательность V1 V2 V3 V4, где V1#V2, V2#V3, V3#V4. 3) Однословный рефрен может быть вставлен в метрически тождественные стихи в начале или в конце стиха или между сегментами, при этом он не составляет части сегмента и не выступает как сегмент:

Udyn', udyn' ur'akaj, mon udokšnyn'.
A — par(o) onne, ur'akaj, mon n'ejekšnyn'
Butto er'ata k, ur'akaj, v'el'(e)ušoso...

Я спал, я спал, невестка, я отсыпался.
Плохой сон, невестка, я видел:
будто бы мы, невестка, живем за деревней...

Предлагаемая система аксиом дает исчисление метрических видов мордовского стиха и их реализаций. Выделяя конструктивные элементы метра и принципы их взаимосвязи, она помогает проникновению в механизм его строения.

Аксиоматическое описание в качестве продолжения предполагает дескриптивное обследование стихового материала с целью проверки и конкретизации обобщенных данных.

Аксиоматический метод нашел плодотворное применение на материале эстонского стиха в исследованиях Я. Р. Пыльдмяэ. Системы стихосложения эстонской поэзии не были достаточно *полно описаны ранее. На основе возможностей фонологической* системы эстонского языка было постулировано существование 8 систем стихосложения. Следующее за этим дескриптивное обследование подтвердило их существование.

Аксиоматическое описание дает возможность не только представить наличное, но и указать на то, что могло бы существовать в *рамках поэзии данного языка*, но в *действительности* не нашло выражения. Оно дает материал для решения вопроса о том, свойственны ли те или иные принципы стихосложения данному языку.

ГЕНЕРАТИВНОЕ СТИХОВЕДЕНИЕ

Генеративное стиховедение, называемое обычно генеративной метрикой, оформилось как направление на материале английского стиха. Сам факт его возникновения и методологические приемы связаны с генеративной лингвистикой. Родоначальниками генеративного направления в стиховедении являются М. Халле и Дж. Кайзер.

Главная задача этого направления состоит в том, чтобы построить лингвистическую теорию, способную предсказать многообразие реализаций метра в языковую форму или свести это многообразие к единству. Эта теория включает в себя три категории составляющих: глубинные, поверхностные и правила корреспонденции.

Основным глубинным компонентом является абстрактная метрическая модель. Она представляет собой упорядоченную цепочку элементов СМ и сл и может содержать дополнительную информацию о них. Назначение правил корреспонденции состоит либо в получении многообразия поверхностных реализаций из глубинной структуры, либо в сведении этого многообразия к ней. В первом случае они должны описать возможные средства и условия лингвистических замен в метрической цепочке, приводящие к реальному многообразию. Во втором — замены, обеспечивающие унификацию. Рассмотрим основные положения теории М. Халле и Дж. Кайзера, построенной в применении к пятистопному ямбу английской поэзии.

Исходным представлением этой теории является абстрактная метрическая модель: (сл)СМслСМслСМслСМслСМ(х)(х), где х может заполняться только безударным слогом, а элементы в скобках могут опускаться. Правила корреспонденции делятся на два типа.

(1) Каждая позиция сл и СМ может быть заполнена одним слогом, либо звуковой последовательностью, содержащей самое большее два гласных (смежных или разделенных одним согласным).

(2) а) Ударные слоги возникают в позициях СМ и во всех позициях СМ; б) ударные слоги возникают только в позициях СМ, но не обязательно во всех; в) ударные максимумы возникают только в позициях СМ, но не обязательно в каждой из них. (Под ударным максимумом понимается ударный слог, расположенный между двумя безударными в пределах одной синтаксической группы и строки.)

Последовательность условий в пределах каждого правила существенна. Каждое следующее условие отменяет какие-либо ограничения предыдущего, но в то же время включает в себя все допускаемые им возможности. В силу этого данная последовательность правил не только охватывает многообразные реализации метра, но и позволяет оценить степень сложности той или иной строки (трудность узнавания метра в ней).

Изложенная теория обобщает значительное многообразие реализаций английского пятистопного ямба. Она может применяться при анализе стиха с целью распознания метра и оценки на его фоне индивидуальных характеристик отдельных строк, в частности, выявления выразительных ходов текста. Она дает также компактное знание о механизме стиха, полезное и для поэта, и для исследователя. Исследователь, напри-

мер, опираясь на круг допускаемых возможностей и степени их сложности, может оценить индивидуальные характеристики произведения. Но самым главным в подобной теории является та обобщенная информация о минимуме условий, которые позволяют отграничить стих данного вида от другого.

Рассмотрим на примерах приложение этой теории к анализу стихов. Строчка

(—) Twenty bookes clad in blak or reed,
сл СМ сл СМ сл СМ сл СМ сл СМ

для носителя английского языка является простым случаем пятистопного ямба. Она удовлетворяет первым условиям правил (1) и (2).

Строки

(—) Twenty bookes clad in blak or reed.
(сл) СМ сл СМ сл СМ сл СМ сл СМ
Are driven like ghosts from an enchanter fleeing
сл СМ сл СМ сл СМ сл СМ слСМ сл

уже сложнее. В первой из них усечен начальный слог. Во второй на СМ стоит безударный слог, что нарушает правило (2, а). Кроме того, первое СМ занимает два слога. Еще более сложной является строка

Wylugh, elm, plane, assh, box, chasteyn, lynde, laurer
сл СМ сл СМ сл СМ сл СМ сл СМ

Здесь имеются ударения на сл, что нарушает правила (2, а) и (2, б), кроме того, нарушено первое условие правила (1). Эта строка, однако, еще соответствует пятистопному ямбу, так как ударения на сл не являются ударными максимумами (2, в). А вот в строке

How meny bards gild, the lapses of time
сл СМ сл СМ сл СМ сл СМ сл СМ

нарушено и последнее условие правил корреспонденции: ударение на предпоследнем сл является еще и ударным максимумом. Эта строка нарушает метр.

Тот факт, что генеративное направление оформилось на материале английского стиха, вызван еще и тем, что английский силлабо-тонический стих дает относительно большое многообразие в варьировании исходной метрической схемы. Большее, например, чем русский. Для четырехстопного ямба русской классической поэзии адекватной является следующая генеративная модель, построенная по аналогии с теорией Халле и Кайзера:

Метрическая схема слСМслСМслСМслСМ (х).

Правила соответствия:

(1) а) Последнее СМ всегда ударно;

(2) а) Ударения возникают на всех СМ и только на них; б) ударения возникают только на СМ, но не обязательно на всех; в) ударение неодносложного слова может падать только на СМ (недопустимость переакцентуации), но не обязательно на каждое из них.

Эти правила проще английских. Здесь нет усечений, возможности вставлять дополнительные слоги, переакцентуировать неодносложные слова.

Собственно генеративная метрика ставит своей задачей провести соответствие между метрической схемой и ее воплощениями с той степенью детализации, которая релевантна признаку метричности. Возможна и бо́льшая степень детализации поверхностных воплощений метра, но она будет захватывать уже не только сферу метрики, но и сферу ритмики.

Генеративное стиховедение не ставит своей задачей описание реальных процессов порождения метрически организованного текста или его восприятия.

РЕКОНСТРУКТИВНОЕ МОДЕЛИРОВАНИЕ СТИХОСЛОЖЕНИЯ

Реконструктивное моделирование стихосложения является новым направлением структурного стиховедения. Оно оформилось в работах М. А. Красноперовой на материале русского стиха. Его основная цель состоит в моделировании внутренней системы, ведающей порождением и восприятием стиха в той мере, в какой это находит выражение в его метрике и ритмике.

Данное направление исходит из того, что ритменный текст обслуживает более или менее развитая специализированная система, имеющая принципиальные черты сходства у разных носителей стихового опыта. В приложении к конкретным метрам ее можно рассматривать как языковую систему метра. Гипотетическое описание языковой системы метра является первой задачей реконструктивного моделирования. Вторая задача состоит в моделировании ритменного текста. Третья предполагает моделирование процессов перехода от глубинных коррелятов ритменного текста к его окончательному текстовому воплощению.

Предположение о сходстве языковой системы метра у разных носителей стихового опыта не равносильно предположению о ее тождестве. Один и тот же естественный язык также имеет принципиальные черты сходства у разных его носителей. Тем не менее и в этом случае нельзя отрицать существования различий. Языковая система метра, как ожидается, менее устойчива, чем естественный язык, поэтому колебания здесь могут быть больше. В связи с этим важно построить достаточно удобную рабочую модель этой системы, чтобы, рассматривая на ее

фоне реальные тексты, можно было бы судить о том, подходит ли она для их описания и какие изменения следует внести для достижения адекватности.

В настоящее время построен рабочий вариант модели порождения и восприятия РС стихотворного текста в применении к классическим размерам русского стиха. Он включает в себя модель ритменного текста и языковой системы метра. Данная модель состоит из двух компонентов — механизма рецепции и механизма генерации.

Механизм рецепции является основной базой ритменного текста. Он осуществляет прием и взаимодействие РС стихотворных строк. Механизм генерации отражает и запоминает процессы, происходящие в сфере ритменного текста, и выполняет порождающие и контролирующие функции. Ядром этого компонента является языковая система метра.

Механизм рецепции состоит из оперативной памяти и механизма реализации РС стихотворной строки. Каждая ритмическая строка, поступающая при восприятии и формируемая при порождении текста, откладывается в оперативной памяти от ее начала. Это соотносится с эффектом регулярного возвращения к началу, который сопровождает обычно восприятие стихотворного текста. Благодаря этому и происходит в модели соизмерение ритмических строк. Механизм реализации осуществляет это наложение. Основными процессами, происходящими в рамках механизма рецепции являются взаимодействие и накопление.

Взаимодействие осуществляется между состоянием оперативной памяти и текущей ритмической строкой. Состояние оперативной памяти определяется при этом тем совокупным коррелятом воспринятых ритмических строк текста, который образовался в ней к данному моменту. В процессе взаимодействия возникают ритмические эффекты. Ритмическим эффектом называется любое изменение и воздействие, которое испытывает оперативная память и сама ритмическая строка, если оно фиксируется языковой системой метра. Например, говоря словами А. Белого, эффект замедления:

Уж осень семьдесят восьмую
Дотягивает старый век...

или эффект ускорения:

Но десять бьет. Он выезжает,
Он полетел. Он у крыльца...

Последовательность ритмических эффектов, вызываемая одной ритмической строкой, называется ритмическим циклом. Последовательность ритмических циклов и пауз между ними, соответствующая целому тексту, называется ритмом текста. Процесс взаимодействия и соответствующий ему ритм называется ритменным текстом.

Совмещение взаимодействия и накопления оказывается возможным благодаря тому, что оперативная память состоит из двух слоев — верхнего и нижнего. Между ними происходит постепенный обмен, направленный на уравнивание выделенностей, оставляемых ритмическими строками. Каждая ритмическая строка поступает непосредственно только на верхний слой, где и осуществляется взаимодействие. Этот процесс может модифицировать состояние верхнего слоя. На нижнем слое взаимодействие происходить не может. Здесь идет постепенное накопление выделенностей.

Процесс накопления призван объяснить, каким образом из строк с разным положением ударений возникает жесткая последовательность СМ и сл, воплощающая метр. В модели это возможно благодаря накоплению выделенностей слогов, происходящему в результате того, что каждая ритмическая строка текста откладывается в оперативной памяти. Выделенные зоны (СМ) возникают в тех ее частях, на которые часто падают ударные слоги, не выделенные (сл) — в тех, на которые они падают редко. Наполняющий СМ и сл коррелят воспринятой последовательности строк называется внутренним метром. Его можно поставить в соответствие тому ударному эталону, который выделяется при скандировании стиха. Можно говорить также, что внутренний метр является модельным отображением профиля ударности.

Механизм генерации включает в себя языковую систему метра и объекты, не относящиеся по своему происхождению к метрике и ритмике. Ядром языковой системы метра в рамках этого компонента является внешний метр. Он состоит из механизма разметки и механизма ипостас. Механизм разметки является динамическим («скандирующим») отображением внутреннего метра. Его функция состоит в том, чтобы обеспечить в оперативной памяти аналог внутреннего метра в случае, когда порождение или восприятие текста только начинается или возобновляется после перерыва.

Механизм ипостас состоит из операторов внешнего метра и управляющих программ. Операторы изначально возникают как отражение воспринятых из текстов ритмических эффектов и процессов, их создающих. Каждый оператор является процедурой осуществления определенного ритмического эффекта в окрестности одного или нескольких соседних СМ. Каждый из них содержит информацию о вероятности соответствующего ему эффекта. Реализация оператора при порождении текста приводит к созданию подходящего эффекта.

Управляющие программы определяют зоны применения оператора в оперативной памяти и его характер. Они ведают также изменениями, происходящими в механизме ипостас, и другими процессами.

Основными функциями внешнего метра являются порождение, контроль и отражение.

Ведущую роль в процессе порождения играет механизм ипостас. Его главное назначение при этом состоит в создании через свои операторы предварительного ритма. Предварительный ритм является той основой, на которую накладывается словесный текст (ср. приводимое во введении высказывание В. В. Маяковского).

Действия внешнего метра в процессе порождения могут подвергаться влиянию корректирующих факторов. Некоторые из них принадлежат языковой системе метра. К их числу относится, например, фильтры и стимуляторы, первые из которых ограничивают, а вторые побуждают использование тех или иных ритмических эффектов. Некоторые корректирующие факторы выходят за пределы моделируемого механизма, выражая более свободные проявления создателя поэтического произведения, в том числе его намерения подчинить ритм семантике текста.

Контроль и отражение осуществляются внешним метром как при восприятии, так и при порождении текста. В обоих случаях это помогает избежать неправильных с точки зрения данной метрической системы вариантов и расширить арсенал используемых средств. При восприятии эти функции внешнего метра способствуют усвоению нового опыта и взаимодействию с уже накопленным. При порождении они помогают предугадать будущую реакцию на ритменный текст, создаваемый поэтом.

Усвоение нового опыта осуществляется, прежде всего, за счет изменения вероятностей и состава операторов механизма ипостас под влиянием воспринимаемых текстов. В процессе порождения этот механизм воспроизводит через операторы накопленный опыт во вновь создаваемых текстах. Согласно модели это является одним из главных условий, обеспечивающих преемственность.

Модель порождения и восприятия охватывает две задачи реконструктивного моделирования. Она содержит свою концепцию ритменного текста и языковой системы метра. Третьей задачей является моделирование процессов перехода от глубинных коррелятов ритменного текста к его окончательному текстовому воплощению.

Исходной рабочей моделью при изучении этой задачи является *модель прозометрической версификации: поэт пишет стих*, как прозаик прозу, отличаясь от него лишь стремлением соблюсти размер. В зависимости от того, как реагирует поэт на несоответствие метру, различают два типа версификации: 1) версификация без перебоев — подходящей считается лишь та строка, которая в процессе формирования не нарушалась инородными для данного метра словами или выражениями; 2) версификация с перебоями — подходящей считается и та строка, в процессе

формирования которой подобные нарушения происходили, но в итоге была найдена соблюдающая метр последовательность.

Для версификации с перебоями вводится понятие количества усилий, отводимых на формирование ритмической строки. В конечном счете оно определяется количеством попыток, которое отводится на поиск подходящего по метру продолжения к уже найденной части строки.

Впоследствии эта исходная теория расширяется. Вводится понятие прозометрической версификации с расщеплением. Стихосложение этого типа уже предполагает более сложный процесс — в частности, возможность отбраковки уже сформированной строки в случае несоответствия ритмики и семантики.

Теория прозометрической версификации, как и изложенная ранее теория порождения и восприятия РС, ориентируется не на идеальное соответствие реальному процессу, а на создание удобной рабочей модели для сопоставления со стихом. Предполагается, что при достаточно удобной фоновой основе текст сам подскажет, каким образом следует модифицировать исходную модель для достижения надлежащей адекватности.

Тем мостом, который соединяет разработанную теорию с текстовой реальностью, являются вероятностные модели размера. Каждая из них показывает, какими могут быть частоты ритмообразующих элементов текста, если выполнен тот или иной комплекс условий стихосложения. Соответствие модели реальным статистическим показателям считается аргументом в пользу данного способа версификации, несоответствие — аргументом против. Различаются модели языкового типа и модели стихового типа. Первые построены на прозаических частотах ритмических слов, вторые — на стиховых частотах ритмических слов. К числу моделей первого типа относится, в частности, языковая модель размера.

Моделям прозаического типа соответствует случай неразвитой языковой системы метра и та или иная вариация прозометрического стихосложения. Моделям стихового типа — случай развитой языковой системы метра. С позиций реконструктивного моделирования ритмическая строка в этом случае складывается не только под действием языка и размера, но может быть задана более или менее подробно той глубинной заготовкой, которую создает механизм ипостас и корректирующие его действия факторы. Формируемое словесное наполнение должно укладываться в эту заготовку.

Описанный аппарат предназначен для гипотетической реконструкции реального стихосложения по статистическим показателям ритмики. Центральную цепочку, по которой осуществлялся процесс реконструкции, можно представить следующим образом: 1) рассматриваются статистические показатели текста; 2) ищется вероятностная модель размера, соответствующая этим показателям; 3) выводится формальная модель стихосложения, со-

ответствующая данной вероятностной в рамках теории реконструктивного моделирования; 4) восстанавливается, по мере возможности, комплекс условий стихосложения, соответствующий этой формальной модели.

С помощью аппарата реконструктивного моделирования изучено несколько этапов развития ритмики русского четырехстопного ямба. Теория прозометрической версификации с определенными видоизменениями оказывается применимой к стиху М. В. Ломоносова. Он написан в период зарождения и начала ритмической эволюции этого размера. Согласно прилагаемой концепции языковая система метра, соответствующая этому стиху, проявляется, главным образом, в действии механизма разметки и фильтров, ограничивающих некоторые виды ритмических строк или слов. Система фильтров при этом усложняется с течением времени. Теория, предусматривающая развитую языковую систему метра, включающую в себя механизм ипостас, применима к более позднему стиху. При достаточной развитости языковой системы метра ее инерция может вызывать противодействие со стороны поэта, что поддерживается, например, показателями по стиху А. Белого.

Полученные данные позволяют предполагать, что при поиске словесного наполнения строки поэты, как правило, заботятся, прежде всего, о том, чтобы найти нужное первое и последнее слово, а середину приспосабливает к ним:

Восторг . . . . . пленил
внезапный

Особенно жестко это проявляется в период зарождения силлабо-тоники в поэзии М. В. Ломоносова.

В рамках системы реконструктивного моделирования получили обоснование и практическую поддержку гипотезы о причинах некоторых универсальных и устойчивых явлений в ритмике русского стиха, в частности, закона регрессивной акцентной диссимиляции и связанных с ним особенностей текстов, написанных четырехстопным ямбом. Проведен эксперимент на ЭВМ по предсказанию ритмических характеристик стиха этого размера в разные периоды XIX в. на основе данных предшествующих периодов. Полученные результаты оказались близкими к реальным.

Основной базой реконструктивного моделирования являются данные, полученные дескриптивным стиховедением. В свою очередь, реконструктивное моделирование выдвигает перед дескриптивным стиховедением свои задачи. Они связаны с проверкой и конкретизацией моделей, с испытанием справедливости их предсказаний.

работ Б. В. Томашевского. Это направление получило заметное развитие в исследованиях А. Н. Колмогорова, А. В. Прохорова и других ученых. Наиболее продвинутой областью вероятностно-статистического метода является теория моделей размера. Их создание начинается с языковой модели. Прообразом этой модели были построения Б. В. Томашевского, окончательный вариант предложен А. Н. Колмогоровым. В основе ее современного варианта лежат три гипотезы о ритмике художественной прозы: 1) об однородности; 2) о независимости; 3) о постоянстве вероятностей в разных позициях текста.

Согласно первой гипотезе в разных произведениях художественной прозы относительные частоты одних и тех же ритмических слов приблизительно равны. Вторая гипотеза утверждает, что вероятность появления того или иного ритмического слова в определенной позиции текста не зависит от того, какие РС его окружают. Согласно третьей гипотезе вероятность появления одного и того же ритмического слова одинакова на разных местах текста.

Все гипотезы проходили проверку на больших выборках из произведений русской художественной прозы и получили удовлетворительное соответствие. Эти результаты дают возможность использовать утверждения, заложенные в гипотезах, в качестве рабочих посылок при создании моделей размера.

Опираясь на эти посылки, вероятность любого ритмического слова в прозе можно оценить его средней частотой в выборках из разных текстов. Вероятность любой РС, состоящей из последовательности ритмических слов, можно вычислить как произведение их вероятностей. Например. В РС вида ◡◡́◡/◡́◡/◡́/◡◡́ вероятности составляющих слов оцениваются числами 0,169; 0,151; 0,063; 0,167. Следовательно, вероятность появления такой РС в прозе будет определяться произведением $0{,}169 \times 0{,}151 \times 0{,}063 \times 0{,}167 = 0{,}0003$. Если прозаические вероятности всех ритмических строк размера рассчитаны, можно получить их стиховые вероятности. Для этого надо сложить полученные числа, а затем поделить каждое из них на общую сумму. Например, прозаическая вероятность всевозможных ритмических строк четырехстопного ямба оценивается числом 0,069. Тогда стиховая вероятность варианта ◡◡́◡/◡́◡/◡́/◡◡́ будет рассчитываться как отношение $0{,}0003/0{,}069 = 0{,}0043$. Рассчитанные таким образом вероятности и составляют языковую модель размера.

МАТЕМАТИЧЕСКОЕ СТИХОВЕДЕНИЕ

Наиболее развитыми направлениями математического стиховедения в метрике и ритмике являются: а) вероятностно-статистическое моделирование и б) автоматизация.

Вероятностно-статистическое моделирование. Вероятностно-статистическое моделирование как направление начинается с

А. В. Прохоровым получен результат, позволяющий предполагать, что в языковой модели отражаются некие базовые отношения, так или иначе трансформируемые в реальном стихе. На материале 120 профилей ударности русского четырехстопного ямба сделаны теоретические расчеты. Их цель состояла в том, чтобы предсказать, какими станут вероятности ритмических форм, если сначала они были такими же, как в языковой модели, а в итоге должны составить задаваемый стихом профиль ударности. Рассчитывалось такое распределение ритмических форм, которое было в данных условиях наиболее вероятным. В подавляющем большинстве случаев теоретические результаты оказались близкими к реальным.

Языковая модель стала важным рабочим инструментом стиховедения, но сама по себе она далеко не всегда оказывается адекватной ритмике текстов. Вопрос о том, какие конкретные трансформации могут переводить базовые отношения в те или иные реальные, является предметом специального изучения. Выясняется, например, что в начале ритмической эволюции четырехстопного ямба частоты ударений на трех первых СМ хоть и меняются, но сохраняют между собой почти ту же пропорцию, которую имеют в языковой модели. Если пересчитать модель на среднюю ударность стиха, получится результат, очень близкий к реальному:

СМ	I	II	III
текст	0,939	0,759	0,525
пересчит. ЯМ	0,945	0,738	0,538

К математическому стиховедению примыкает и тот аппарат моделей размера, который разработан в рамках реконструктивного моделирования. Модели, опирающиеся на прозаические вероятности, но не относящиеся к описанной языковой, можно рассматривать как более или менее сложные трансформации, полученные путем видоизменения заложенных в ней принципов. Наряду с данной языковой среди них выделяется другая базовая модель, которая, в отличие от нее, построена по принципу зависимости ритмических слов, следующих друг за другом. Путем изменения порядка подчинения слов, видоизменения их вероятностей, комбинирования принципа зависимости и независимости построен ряд новых моделей, некоторые из которых оказались более адекватными текстам, чем языковая. Все они вместе с языковой моделью связаны единой вероятностной концепцией последовательности повторных независимых испытаний. Разным моделям ставятся в соответствие лишь разные системы событий, полученные на базе этой последовательности. Это дает возможность сравнивать между собой принципы строения разных текстов, выделяя то общее, что их объединяет.

Модели стихового типа в качестве вероятностей ритмических слов используют их средние частоты в соответствующих пози-

циях стихотворных текстов. Вероятность слова в них всегда зависит от его места в строке и может зависеть или не зависеть от того, что ему предшествует.

Тот факт, что все модели размера опираются на ту или иную статистику ритмических слов, делает вопрос о ее получении одной из важных вспомогательных задач стиховедения. Математическое стиховедение разрабатывает теоретические способы вычисления вероятностей ритмических слов для того, чтобы повысить объективность и гибкость расчетов. В первую очередь это касается частот ритмических слов в прозе.

Методика, предложенная А. Н. Колмогоровым и А. В. Прохоровым, различает три статистики. Первая учитывает почти все слова от пробела до пробела. Это статистика элементарных слов. Вторая статистика получается в результате присоединения в качестве энклитик и проклитик предлогов и других несамостоятельных слов (*у реки, за домом* и т. д.). Третья определяется после аналогичного атонирования всех двойственных слов (слов, которые могут быть как ударными, так и безударными: местоимений, вспомогательных глаголов и т. д.). Первая статистика подсчитывается по текстам, вторая и третья выводятся теоретически. Теоретические данные сравнивались с показателями «Пиковой дамы» А. С. Пушкина и «Бэлы» М. Ю. Лермонтова и обнаружили хорошее приближение.

Математическое стиховедение занимается не только теоретическим моделированием, оно разрабатывает также способы непосредственного описания текстов. Примером может служить комплексная характеристика В. С. Баевского, названная им показателем изменчивости ритма. Каждому слогу сопоставляется число, называемое мерой выделенности. Для каждой слоговой позиции вычисляется средняя выделенность по всему произведению. Таким образом получается усредненная строка, представляющая собой последовательность мер выделенности слогов. Она служит фоном, на котором рассчитывается ритмическая характеристика реальной строки. Эта характеристика определяется как среднее квадратичное отклонение данной строки от соответствующего ей фона. Средняя ритмическая характеристика для всех строк и называется показателем изменчивости ритма.

Автоматизация. Применение ЭВМ в стиховедении имеет более чем двадцатипятилетнюю историю. На материале русского стиха это направление работ было начато в исследованиях М. А. Красноперовой, В. С. Баевского и др. Несмотря на разнообразие исследовательских целей в области метрики и ритмики сама машина использовалась главным образом для выполнения статистических подсчетов по текстам и расчета моделей. В последнее десятилетие на ЭВМ стали выполняться и относительно простые стиховедческие задачи логического характера. Это, в первую очередь, разработки О. П. Исакова по авто-

матическому распознаванию метрической организации текстов, написанных классическими размерами.

В современных условиях встает вопрос о создании диалоговых систем по стиховедению на персональных ЭВМ. Эти системы должны работать в режимах исследования, накопления и обучения. В первом режиме машина должна давать ответы на вопросы по метрике, ритмике и другим, интересующим стиховеда, признакам текстов, во втором — получать сведения о вновь поступающих текстах и вводить их в имеющуюся базу данных, в третьем — обучать стиховедческому анализу произведения.

ПРИКЛАДНОЕ СТИХОВЕДЕНИЕ

Прикладное стиховедение может обслуживать литературоведческие, лингвистические и другие задачи.

В области литературоведения значение стиховедческих исследований обусловлено тем, что они являются частью поэтики — науки о строении литературного произведения. Цели, выдвигаемые поэтикой перед стиховедением, связаны, с одной стороны, с изучением жанровых и стилистических типов, с другой стороны — с изучением свойств отдельных произведений, творчества определенного автора, школы, литературной эпохи и т. д.

Стиховедение располагает сейчас значительными данными о метрических и ритмических тяготениях жанровых категорий (лирика, эпос), их выраженной зависимости от литературного направления (классицизм, романтизм, реализм) или автора.

При изучении отдельного произведения стиховедческий анализ может применяться для более глубокого проникновения в его содержание или другие сферы нестиховедческого характера. Характерной канонической формой описания авторской манеры стали метрические справочники поэтов. В справочниках отражается общий метрический репертуар поэта, его статистические характеристики, изменение во времени, метрические характеристики отдельных жанров, произведений и другие сведения. Прикладное стиховедение располагает сейчас справочниками по произведениям А. С. Пушкина, М. Ю. Лермонтова, Е. А. Баратынского, К. Батюшкова и других поэтов. Эти справочники имеют многоцелевое назначение. Они могут использоваться при изучении вклада поэта в развитие стиховой формы, эволюции его творчества, влияния на других авторов. Они дают фоновый материал при анализе отдельного произведения, исходные данные при изучении стиховой манеры поэтической школы, литературной эпохи и т. д. Стиховедение располагает сейчас и возможностями создания ритмических справочников. Эти возможности могут быть значительно расширены с помощью ЭВМ.

Одной из особых задач литературоведения является атрибуция и датировка текстов. Возможности стиховедения в этом

отношении весьма существенны. Стиховедческие результаты при уточнении авторства и даты использовались как в русском, так и в иноязычном литературоведении. На материале русского стиха стиховедческий анализ помог установить, например, что автором «Гавриилиады» является А. С. Пушкин, а продолжение «Русалки», предложенное Д. П. Зуевым, противоречит показателям пушкинской ритмики.

Прикладное стиховедение находит применение в поэзии и художественном переводе. Переломные этапы в истории русской поэзии всегда были связаны с деятельностью поэтов — исследователей и экспериментаторов в области стихотворной формы. С нуждами поэзии связан сам факт возникновения стиховедения как науки. В период зарождения и становления силлабо-тоники (вторая треть XVIII в.) проблемы стихотворной формы и, в первую очередь, метра и ритма были предметом изучения, дискуссий и нормативных установок ведущих поэтов того времени — М. В. Ломоносова, В. К. Тредиаковского и А. П. Сумарокова.

Серебряный век русской поэзии стал и веком зарождения русского стиховедения в его современном виде. Его основатель, А. Белый, был в то же время одним из наиболее крупных поэтов-символистов. Значительное внимание изучению стихотворной формы уделяли В. Я. Брюсов, Н. С. Гумилев и другие видные поэты. Главные стиховедческие труды Брюсова трактуются современным исследователем (С. И. Гиндин) как прообраз генеративной метрики.

Известен интерес с поэтической форме А. С. Пушкина. Наиболее крупным исследователем стиха в пушкинское время был поэт и переводчик А. Х. Востоков. Влияние его стиховедческих разработок сказалось на ритмике «Песен западных славян» Пушкина.

Стихотворный перевод также является одним из источников, вызывающих к жизни прикладное стиховедение. Даже близкие разноязычные системы стихосложения различаются особенностями в деталях. Вспомним, например, различия русского и английского ямба. Во многих случаях арсенал стихотворных форм настолько различен, что без тщательного изучения принципов и практики иноязычного стихосложения становится невозможным и качественный перевод. Совмещение переводческой и стиховедческой деятельности вполне обычно. Создатель классического русского перевода «Илиады» Гомера Н. И. Гнедич был одновременно и создателем классической формы русского гекзаметра. Сам он опирался на гекзаметр, созданный в переводческой работе В. К. Тредиаковского. Переводчиками являются также крупные стиховеды нашего времени — М. Л. Гаспаров, А. Л. Жовтис и др.

Задачи стиховедения в области лингвистики связаны с тем, что стихотворные тексты представляют собой речь особого ро-

да. В силу этого лингвистический интерес имеют любые результаты их стиховедческого анализа. Лингвистика дает лишь свое направление в осмыслении этих результатов. Один из аспектов этого направления касается взаимоотражения стиховедческих и общелингвистических категорий — грамматики, фонетики, стилистики. Сюда относятся такие проблемы, как ритм и синтаксис, ритм и интонация, стилистические типы ритма и т. п. Самодовлеющую роль в ритмическом строении стиха играет его слоговая организация. Изучение свойств слоговых структур может быть продуктивным направлением прикладного стиховедения.

Важный аспект лингвистических приложений связан с осмыслением ритменного текста как особой коммуникативной формы, а метра — как языка, ее порождающего. Ритменный текст относится к тем исключительным феноменам, в которых находят формальное выражение коммуникативные формы, предшествующие вербальному тексту. В связи с этим на материале ритменного текста можно ставить задачу изучения дословесных языковых форм. Эта задача лежит на грани лингвистики и психологии.

Потенциальная связь стиховедения с психологией проявляется не только в сугубо лингвистических задачах. Прикладные возможности этой науки в данном случае обусловлены тем, что она имеет дело с формальным воплощением явлений, фигурирующих в сфере бессознательного. Языковую систему метра можно рассматривать как стихийно возникающий самообучающийся механизм, способный порождать и распознавать образы. Ее можно изучать поэтому как пример механизмов такого рода. Стиховедение способно обслуживать некоторые специфические проблемы и в области психологии творчества. Основной вопрос здесь касается взаимодействия канонизированной знаковой системы и свободных проявлений поэтического настроя.

Прикладное стиховедение имеет свои задачи и в сфере искусственного интеллекта. Их назначение состоит, во-первых, в моделировании на ЭВМ определенных функций человека при порождении и восприятии стиха: а) распознавание метра и эквивалентных форм организации; б) реакция на ритмические ходы текста; в) самообучение и воспроизведение метрической системы. Во-вторых, это могут быть модели диахронического и прогнозирующего характера: а) воссоздание на ЭВМ истории перехода одних ритмических и метрических типов в другие; б) прогнозирование их будущего развития. В-третьих, это может быть реализация стиховедческих возможностей при машинном переводе и обучении, например, автоматический подбор подходящей метрической системы, обучение анализу стихотворного текста.

Таким образом, стиховедение имеет разнообразные прило-

жения, многие из которых стимулируют перспективные разработки.

Литература

Баевский В. С. Стих русской советской поэзии: Пособие для слушателей спецкурса. Смоленск, 1972. 145 с.

Белый А. Символизм. М., 1910. С. 231—395.

Гаспаров М. Л. 1) Современный русский стих: Метрика и ритмика. М., 1974. 487 с.; 2) Очерк истории русского стиха: Метрика. Ритмика. Рифма. Строфика. М., 1984. 319 с.; 3) Очерк истории европейского стиха. М., 1989. 303 с.; 4) Русские стихи 1890—1925 годов в комментариях. М., 1993. 372 с.

Гиндин С. И. Трансформационный анализ и метрика: (Из истории проблемы) // Машинный перевод и прикладная лингвистика. Вып. 13. М., 1970. С. 177—200.

Жирмунский В. М. Введение в метрику // Жирмунский В. Теория стиха. Л., 1975. С. 5—232.

Колмогоров А. Н., Прохоров А. В. Модель ритмического строения русской речи, приспособленная к изучению метрики классического русского стиха: (Введение) // Русское стихосложение: Традиции и проблемы развития. М., 1985. С. 113—133.

Краснопёрова М. А. 1) Метр и ритм стихотворного текста: (Модель) // Сборник научных трудов / МГИИЯ им. Мориса Тореза. Вып. 307: Проблемы поэтической речи. М., 1988. С. 91—96; 2) Модели лингвистической поэтики. Ритмика: Учебное пособие. Л., 1989. 87 с.; 3) Об автоматическом анализе ритмики русского стиха (обзор) // Машинный фонд русского языка: Предпроектные исследования. М., 1988. С. 242—267; 4) О глубинных единицах стиховой ритмики // Сборник научных трудов / МГИИЯ им. Мориса Тореза. Вып. 352: Единицы поэтической речи. М., 1990. С. 13—26.

Прохоров А. В. О случайной версификации: (К вопросу о теоретических и речевых моделях стихотворной речи) // Проблемы теории стиха. Л., 1984. С. 89—98.

Руднев П. А. Введение в науку о русском стихе. Тарту, 1989. 121 с.

Тарановски К. 1) Руски дводелни ритмови. Београд, 1953. 376 с.; 2) О ритмической структуре русских двусложных размеров // Поэтика и стилистика русской литературы. Л., 1971. С. 420—428.

Тарлинская М. Г. Применение точных методов в стиховедении для целей атрибуции (на опыте изучения драм Шекспира) // НТР и развитие художественного творчества. Л., 1980. С. 175—184.

Теория стиха / Отв. ред. В. Е. Холшевников. Л., 1968. 316 с.

Томашевский Б. В. 1) О стихе: Статьи. Л., 1929. 326 с.; 2) Стих и язык: Филологические очерки. М.; Л., 1959. 471 с.

Трубецкой Н. С. О метрике частушки // Избр. труды по фонологии. М., 1987. С. 371—390.

Тынянов Ю. Н. Проблема стихотворного языка // Тынянов Ю. Н. Проблема стихотворного языка. М., 1965. С. 21—194.

Халле М., Кайзер С. Дж. Английский пятистопный ямб // Новое в зарубежной лингвистике. Вып. 9. М., 1980. С. 356—378.

Холшевников В. Е. Основы стиховедения. Русское стихосложение: Пособие для студентов филол. фак. 2-е изд., перераб. Л., 1972. 168 с.

Шенгели Г. А. 1) Трактат о русском стихе: Органическая метрика. М.; Пг., 1923. 184 с.; 2) Техника стиха. М., 1960. 312 с.

Якобсон Р. Аксиомы системы стихосложения (на примере мордовской народной песни) / Пер. с англ. Н. В. Перцова // Якобсон, Роман. Работы по поэтике. М., 1987. С. 133—139.

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ДРЕВНИХ ТЕКСТОВ

Автоматическая обработка древних рукописей и подготовка их к последующему изданию имеет большое историко-культурное значение. Немало рукописей на восточных, германских, романских, славянских языках дошли до нас из глубины веков. Изучением древних рукописей занимаются специалисты самых разных профессий: филологи, историки, этнографы, географы, экономисты, социологи.

С точки зрения лингвистов целью подлинно научной истории языка является строгое объективное воссоздание системы языка текстов (текста).

Синхроническое моделирование систем языка древних текстов предусматривает построение различных моделей на уровне морфологии, лексики. Такие модели могут быть как собственно морфологическими, ареально-типологическими, функционально-семантическими, так и хронологически-статистическими, описательными, формальными. При этом следует различать исследования, направленные, с одной стороны, на построение системы языка как таковой, а с другой — на воссоздание истории текста. При моделировании системы языка главным является выявление основных доминирующих типов и тенденций; здесь наиболее показательны морфология, высокочастотные факты в статистике. История памятника связана с проблемами атрибуции текста; здесь главное — выделение диагностических различительных признаков текста. Наиболее показательны в этом отношении письмо текста, его лексика, синтаксис, собственно стилистика, редкие факты статистики.

В процессе моделирования языка особое место занимают словоуказатели. Словоуказатели к древним памятникам литературы нужны языковедам, литературоведам, историкам, этнографам. Они дают полное представление о лексике памятника, о типах слов и форм по их структуре, о характере применения

титл, о видах использования графем и их последовательностях. Словоуказатели демонстрируют абсолютную частоту словоформы в тексте и в целом позволяют судить об общем семантическом характере приведенных слов. Идя по пути создания полных словоуказателей, реестров грамматических форм, конкордансов и словарей к отдельным памятникам (центрам письменности), сопоставляя затем между собой словники и словари текстов и зон, изученных лучше других, мы придем постепенно к относительно полной картине ареального и жанрового распределения лексики памятников письменности и сможем более уверенно говорить о неологизмах изводов, школ и авторов отдельных текстов.

Независимо от задачи, стоящей перед исследователем, вводу текстов большого объема в память ЭВМ должна предшествовать определенная подготовительная работа. Во-первых, текст внимательно изучается с тем, чтобы выявить все типы явлений, которые должна уметь обработать будущая программа. Во-вторых, вырабатывается подробная концепция программы. В-третьих, исходя из этой концепции формулируются правила ввода текста, которые, с одной стороны, не должны вызывать затруднений у работника, вводящего текст, а с другой стороны, не должны требовать громоздких программистских решений. Без двух первых этапов практически невозможно составить оптимальные правила ввода текста, а любое изменение правил ввода в процессе написания программы повлечет за собой корректировку уже введенного текста — задачу тем более неприятную, чем больше его объем. Таким образом, правило «семь раз отмерь — один отрежь» как нельзя лучше подходит к задачам машинной обработки больших текстовых массивов.

При вводе древнерусских текстов приходится сталкиваться с некоторыми специфическими трудностями, отсутствующими при вводе современных текстов. Так, практически невозможно использовать сканер — устройство оптического ввода, так как: а) текст, вообще говоря, нелинеен: имеются выносные буквы и надстрочные знаки; б) встречаются лигатуры — слитные написания нескольких букв; в) в тексте рукописей нет деления на слова вообще, либо выделяются только значащие слова, а служебные пишутся слитно с окружающими словами. Для проверки введенного текста нельзя использовать спеллер — программу выявления опечаток и орфографических ошибок, так как каждая словоформа обычно встречается в нескольких орфографических вариантах и даже в одной строке одно и то же слово может быть написано разными способами.

Но вот текст введен. Теперь предстоит, как это ни странно, почти равная по трудоемкости работа — его выверка. Текст должен проверяться по двум параметрам: во-первых, проверяется тождественность введенного текста оригиналу, во-вторых, проверяется строгое соблюдение всех формальных правил ввода,

так как любое отступление от них отрицательно скажется на работе программы. Отсутствие твердых орфографических норм в древнерусских текстах требует посимвольного сравнения введенного текста с исходным, что в общем-то исключает возможность корректировки текста непосредственно на экране дисплея. Текст, как правило, приходится распечатывать, отмечать ошибки в распечатке и лишь затем вносить исправления в машинную версию текста. При большом количестве ошибок проверку приходится повторять, так как при корректировке не исключено внесение новых ошибок.

Какие задачи машинной обработки древнерусского текста могут стоять перед исследователем? Обычно это задачи составления разного рода словарей — алфавитных, частотных (слова упорядочиваются по частоте), алфавитно-частотных (слова упорядочиваются по алфавиту и для каждого приводится его частота в тексте), словоуказателей или индексов (слова располагаются по алфавиту и для каждого указываются его адреса в тексте — номера листов и, возможно, строк), конкордансов (для каждого слова приводятся все контексты, в которых оно встретилось) и т. д.

Любая программа составления словарей принципиально состоит из трех частей: 1) ввод текста в оперативную память, выделение и накапливание исследуемых единиц текста (например, словоформ) и определение их параметров (например, частот или адресов); 2) сортировка выделенных единиц по избранному параметру (например, по алфавиту или частоте); 3) вывод результата (словаря) на экран дисплея, на принтер или на запоминающее устройство.

При написании программы блока ввода текста в оперативную память полезно обеспечить его независимость от алфавита вводимого источника. Для этого в программе можно предусмотреть обработку лишь типов символов, встречающихся в тексте, а сами символы, т. е. алфавит, вводить в потоке данных, а не задавать жестко в программе. Например, могут быть выделены следующие типы символов: буквы; символы, входящие в состав слова, но не имеющие определенного места в алфавите (например, дефис, апостроф); знаки препинания и т. п.

Сортировка выделенных единиц текста может быть отдельным блоком, но может проводиться и одновременно с вводом. Часто можно использовать стандартные процедуры сортировки, но иногда может возникнуть потребность составить программу сортировки самостоятельно, например, для уменьшения объема используемой памяти.

При написании программ, рассчитанных на обработку текстов значительного объема, необходимо предусмотреть использование долговременной памяти, так как объем оперативной памяти ограничен.

Актуальной нерешенной проблемой в сфере составления ав-

томатических словарей является проблема автоматического сведения словоформ к канонической (словарной) форме. Не говоря уже о трудности определения грамматической характеристики словоформы, следует отметить проблему отождествления графических вариантов одной и той же формы слова. Предпринимавшиеся в этой области усилия пока плодов не принесли, и единственной надежной процедурой сведения словоформ к канонической форме, к сожалению, является ручное постредактирование полученного словаря.

Приведем в качестве примера опыт разработки автоматизированного банка данных по памятникам русской агиографии XV—XVI вв.

Основными жанрами древнерусской литературы являются следующие: повествовательный («Сказание о Мамаевом побоище», «Задонщина» и другие воинские повести), конфессиональный (или конфессионально-повествовательный), в котором особого внимания заслуживает агиография — жизнеописания (жития) святых типа «Жития Сергия Радонежского».

Классическая славистика, и прежде всего русская филологическая школа, сделала немало для выяснения и описания произведений древнеславянской литературы конфессионального жанра. Так, только в агиографии мы располагаем капитальными исследованиями В. О. Ключевского, Н. И. Серебрянского, П. Сырку, Л. А. Дмитриева. К сожалению, позднее эта традиция почти оборвалась. Другие. весьма важные сами по себе проблемы надолго оттеснили на второй план вопрос об изучении того подлинно литературного языка, который функционировал в средние века в письменности восточных и южных славян.

Разработка темы была начата на базе исследований по автоматизации обработки древнерусских текстов, проводившихся с 1980 г. на кафедре математической лингвистики ЛГУ. В ней принимали участие сотрудники кафедры русского языка, а также ИРЛИ АН СССР и ГПБ им. М. Е. Салтыкова-Щедрина.

Решено было обратиться пока к памятникам одного жанра — агиографического (жития святых русской церкви и похвальные слова им XV—XVI вв.), так как эти памятники, с одной стороны, являются наименее изученными в языковом отношении (материал их практически не отражен в «Словаре русского языка XI—XVII вв.»), а с другой стороны, представляют особый интерес для изучения истории собственно литературного языка. Именно в это время и в этом жанре складывались фразеология, стандартные словесные комплексы, которые мастера слова творчески перерабатывали, используя в текстах, построенных по определенным законам жанра. Язык агиографических произведений во многом обусловил судьбу и характер русского языка этого времени. Кроме того, создание подобного банка данных имеет в перспективе большое значение

для изучения истории литературы и церковнославянского языка XV—XVI вв. не только на Руси, но и в Болгарии, Сербии, Словении.

В целях последующей обработки данных рукописей на ЭВМ прежде всего была составлена (в течение 1982—1983 гг.) картотека канонизированных святых русской церкви с указанием реестра их житий, похвальных слов, сказаний, учтены разные издания этих текстов. Картотека базируется на материале исследований Н. П. Барсукова, В. О. Ключевского, М. К. Соколова, А. П. Кадлубовского, Филарета, И. Яхонтова, П. М. Строева, картотек Российской национальной библиотеки. Картотека составлена на строгой основе; к отбору текстов предъявляются определенные требования и прежде всего — требование полноты текста. При этом, когда из нескольких списков одного памятника выбирался один, необходимо было учитывать и экстралингвистические (технические) факторы.

Агиографический фонд кафедры математической лингвистики в настоящее время насчитывает 79 текстов (жития, похвальные слова, памяти и т. д.) XV—XVII вв. Тексты фото- или ксерокопированы. Сейчас в работе находится 60 текстов на разных стадиях обработки.

Словоуказатель в готовом виде представляет собой полный реестр всех слов и словоформ одного или нескольких слов, встречающихся в тексте-источнике, с адресом словоформы по всем введенным в эту систему памятникам.

В дальнейшем информационно-поисковая система запросно-ответного типа позволит выдавать контексты запрашиваемой словоформы, текст запрашиваемого памятника.

Разрабатываемая информационно-поисковая система запросно-ответного типа предоставляет пользователю широкие возможности для работы в режиме диалога. Он сможет узнать реестр всех слов и словоформ одного или нескольких слов; адрес словоформы по всем введенным в эту систему памятникам; все грамматические формы слова, а также орфографические нормы слова; контексты запрашиваемой словоформы, т. е. все типы словосочетаний; обратный словарь словоформ, дающий возможность получить информацию об аффиксах слова; текст запрашиваемого памятника.

Пользователь системы в ходе диалога указывает слова (формы), которые его интересуют, ЭВМ выдает на дисплей для этих слов все имеющиеся контексты и все адреса этого слова в данном тексте.

ДЕШИФРОВКА

Дешифровка — это исследование сообщений или текстов для обнаружения информации, представленной способом, не известным исследователю.

Из приведенного определения следует, что объектом дешифровки можно считать только то, что было написано, начерчено, высечено или вообще как-либо материально зафиксировано человеком с целью передать информацию адресату. Адресатами могут быть современники, потомки, мифические существа, в частном случае сам автор сообщения. Ни изучение нуклеотидных цепей, ни медицинская диагностика, имеющие природные объекты, ни работа детектива, расследующего преступление, т. е. результат человеческой, но не коммуникативной деятельности, не являются, с нашей точки зрения, дешифровкой (хотя некоторые авторы и используют этот термин столь расширительно).

Объектами дешифровки являются тексты на языках каких-либо народов или записи на иных знаковых системах (ноты, картографические документы, математические формулы и т. п.). Этим определяется различие между лингвистической и нелингвистической дешифровкой. Последняя лежит за пределами языковедческих дисциплин, но ее роднят с ними общие семиотические закономерности.

Лингвистическая дешифровка включает в себя исследование исторических письмен и чтение тайнописи (криптограмм). В первом случае непонятность текста для исследователя обусловлена разрывом традиции, во втором — специальными усилиями, предпринятыми автором или шифровальщиком, чтобы сузить круг адресатов.

По мнению Ю. В. Кнорозова, дешифровка исторических систем письма и дешифровка секретных шифров не имеют почти ничего общего. Он прав в отношении тайнописи, основанной

на перестановках, т. е. шифрах, построенных на том, что обычные языковые элементы расставляются в ином, условном порядке. Такого рода шифры действительно имеют мало сходства с записью человеческой речи. Что касается текстов, записанных шифрами и кодами, основанными на замене элементов текста другими, заранее условленными, то в них так же сохраняются естественные закономерности построения, как они присутствуют и в текстах на неизвестных языках, поскольку «во всех языках всех времен используются сходные способы передачи информации» (Кнорозов, 1982, с. 5). В некоторой степени не только шифровка и дешифровка, но даже традиционный перевод может рассматриваться как перекодировка, основанная на замене элементов текста, принадлежащих одной системе кодов, элементами, принадлежащими другой системе.

Поэтому мы склонны присоединиться к П. Аальто, который считает, что между дешифровкой исторических письмен и чтением криптограмм имеется много точек соприкосновения. Именно на их аналогии он строит свою классификацию дешифровочных задач, в основном разделяемую и И. Фридрихом. Но прежде чем рассмотреть ее, надо сделать несколько предварительных замечаний.

Слова «шифр» и «код» часто используются как синонимичные, но в криптологии (дисциплине, изучающей тайнопись) они представляют собой термины и четко противопоставлены друг другу.

Шифр имеет в качестве «означаемого», или номинанта, звук или элемент письменности данного языка. Например, простейшим шифром является обозначение каждой буквы алфавита ее порядковым номером. Означающее в этом шифре — двузначное число, номинант — буква русского алфавита, порядковый номер которой обозначен этим числом.

а—01	и—10	с—19	ъ—28
б—02	й—11	т—20	ы—29
в—03	к—12	у—21	ь—30
г—04	л—13	ф—22	э—31
д—05	м—14	х—23	ю—32
е—06	н—15	ц—24	я—33
ё—07	о—16	ч—25	
ж—08	п—17	ш—26	
з—09	р—18	щ—27	

Известная стихотворная строка «Я помню чудное мгновенье» зашифровывается следующим образом: 33 1716141532 252105151606 140415160306153006.

Несколько хитроумнее так называемый «шифр Цезаря». По сообщению Светония, Гай Юлий Цезарь «пользовался тайнописью, то есть менял буквы так, чтобы из них не складывалось ни одного слова. Чтобы разобрать и прочитать их, нужно читать всякий раз четвертую букву вместо первой, например, d

вместо а и так далее» (Светоний. Божественный Юлий, 56, 6). Приведем таблицу для латинского алфавита эпохи Цезаря:

a—d	l—m	r—u (v)
b—e	k—n	s—x
c—f	l—o	t—y
d—g	m—p	u (v)—z
e—h	n—q	x—a
f—i	o—r	y—b
g—k	p—s	z—c
h—l	q—t	

Имя Gaius Iulius Caesar в этом шифре выглядит так: kdmzx mzomzx fdhxdu.

В отличие от шифра, код имеет означаемое в прямом смысле, без кавычек, т. е. представляет собой условное означаемое для существующего в кодируемом языке значения. Примером может служить «офенский язык», жаргон бродячих торговцев центральной России XIX—XX вв. «Обтыривай маз! Дулец-то яманный!» — говорил коробейник своему напарнику, и тот понимал: «Обвешивай, приятель! Мужик-то простоват!» Обратим внимание на то, что этот код не был искусственно создан конкретным автором, а развился как социальный диалект русского языка. На его языковую принадлежность указывает то, что кодированию, т. е. замене, подверглись только элементы, несущие лексическое значение, а грамматические и словообразовательные морфемы остались обычные, русские. Это позволило ему легко употребляться в устной речи. Коды, обслуживающие письменную речь, могут предполагать замену экспонентов также и служебных морфем. Таким образом, систему письма можно рассматривать как систему шифров, а лексику языка — как множество кодов или систему кодов.

Надо заметить, что сложившийся терминологический узус не совсем удачен, так как шифром или кодом называют: а) отдельный шифр или код; б) систему шифров или кодов или формулу, по которой они созданы; в) текст, записанный в такой системе.

Некоторую неустойчивость терминологии также приходится иметь в виду, обращаясь к типам дешифровочных задач. Лингвистическая дешифровка в широком смысле слова включает в себя решение задач четырех типов.

1. Неизвестна только письменность, язык известен. Аналог: простой шифр, основанный на замене. Решение этой задачи, т. е. установление чтения забытых знаков, есть дешифровка в узком смысле слова.

2. Неизвестен только язык, письменность известна. Аналог: код. Решение этой задачи, т. е. установление значения единиц языка, звучание которых известно, называется интерпретацией.

3. Неизвестный язык записан неизвестным письмом. Аналог:

зашифрованный код. Решение третьей задачи, т. е. установление и звучания, и значения языковых единиц, называется р а с к р ы т и е м.

4. Чтение, перевод и анализ особенностей конкретного текста, записанного на неизвестном языке или неизвестной письменностью. Хотя такая деятельность, опираясь на решение приведенных выше задач, относится уже к области филологии, но она качественно отлична от анализа текста на хорошо известном языке, поэтому ее называют д е ш и ф р о в к о й т е к с т а.

Впрочем, в истории случалось, что не дешифровка конкретных памятников опиралась на решение системных задач, а, наоборот, систему языка или систему письма удавалось распознать благодаря удачному исследованию определенного текста.

Остановимся подробнее на рассмотренных выше типах.

Представим себе текст на русском языке, записанный латинскими буквами. Хотя для передачи некоторых русских букв, например *ч* и *щ*, требуются надстрочные знаки и иные ухищрения, большинство русских, изучавших в школе один из западных языков, сумеют без особого труда такой текст прочитать и понять. Эмоциональная реакция на непривычный облик русских слов относится к социальной и историко-культурной сфере, которую, безусловно, тоже должен учитывать дешифровщик, но здесь мы оставляем ее за скобками и иллюстрируем только семиотический аспект. Дано множество фонем языка, оно может быть передано и русским, и латинским алфавитом, оба из которых нам известны. Несложно научиться переводу записи из одного алфавита в другой, эта работа называется транслитерацией. Такой процедуре, в частности, подвергаются названия всех российских судов, плавающих за пределами территориальных вод. Существует официально утвержденная таблица соответствия русских букв латинским.

Теперь представим себе текст на русском языке, записанный армянскими буквами. (Те, кто знает армянский алфавит, могут представить себе русский текст, написанный каким-то иным письмом.) Перед нами образец первой задачи: неизвестна письменность, язык известен. Более того, это наш родной язык. Но, не умея прочитать, мы едва ли отличим эту запись от записей на армянском языке, которого не знаем. Таким образом, разница между первой и третьей задачей не очевидна, она выявляется только в ходе дешифровки. Конечно, мы можем открыть справочник и выяснить, как читать армянские буквы. Но наше личное незнание армянского алфавита в данном примере лишь моделирует ситуацию, когда мировой науке не известно, каково звучание знаков письма, когда таблицу звуковых соответствий, т. е. номинантов, которую мы знаем для латинского алфавита и можем узнать для армянского, еще не знает никто.

Именно с такими трудностями столкнулись дешифровщики крито-микенского линейного письма Б. Линейное Б — силлаба-

рий, т. е. письмо, в котором номинантом большинства знаков является слог. Начало дешифровке положила А. Кобер, составившая таблицу слоговых знаков, в которой в одной строке оказывались слоги, содержащие одинаковый согласный, а в одном столбце — слоги, содержащие одинаковый гласный. Расширив и дополнив таблицу, М. Вентрис стал подставлять в нее слоги разных языков, в том числе и греческого, который первоначально казался ему самому маловероятным претендентом. Пришлось проделать большую работу, прежде чем ему и его коллеге Дж. Чедуику удалось выяснить, а затем и доказать, что глиняные таблички написаны на архаичном диалекте хорошо известного древнегреческого языка.

Примером второй задачи может служить современный текст на узбекском языке. Те, кто знает узбекский, могут взять в качестве примера тексты на таджикском, осетинском или любом другом языке, пользующемся русской графикой. Читать их, зная русский, нетрудно, т. е. приблизительное звучание можно воспроизвести. Но нельзя понять прочитанное. Известна письменность, неизвестен язык.

Одной из самых знаменитых дешифровок, относящихся к этому типу, является интерпретация хеттского языка, проведенная Б. Грозным. Хеттские тексты написаны клинописью, близкой к аккадской, которая к началу XX в. была уже неплохо изучена. Б. Грозный применил этимологический метод дешифровки, предположив родство хеттского с индоевропейскими языками и найдя соответствующие друг другу фонетически и семантически корни (первым был корень water со значением "вода"). При работе по этимологическому методу дешифровщик строит таблицы звуковых соответствий не знаков письменности, а экспонентов родственных слов и морфем. Нужно иметь в виду, что соответствие не означает обязательного тождества или даже близкого сходства звучания. Например, симпатичное домашнее животное, именуемое по-русски *кот*, на близкородственном украинском называется *кіт*. Выявление звуковых соответствий неизвестного языка и родственных ему — это нелегкий и кропотливый труд. Но это благодарный труд и счастливые обстоятельства. Всю меру их удачности могут оценить лишь те, кто занимается дешифровкой в условиях отсутствия достоверно родственных языков. Так обстоит дело, например, с этрусским языком.

Этрусский алфавит родствен греческому и латинскому. Работа по отождествлению разных написаний одних и тех же букв завершена, и чтение знаков этрусского письма практически не представляет труда. Но пример с узбекским текстом должен напомнить нам, что чтение и понимание — это разные вещи.

При интерпретации этрусского языка использовались два основных метода: этимологический, который мы уже рассматривали в связи с хеттским, и комбинаторный.

Этимологический метод основывается на поисках языка, родственного этрусскому. В качестве кандидатов на родство с ним привлекалось множество языков от баскского на западе до японского на востоке, от германских и балто-славянских на севере до дравидийских на юге. Сложность заключается в том, что родство языков может быть доказано путем установления закономерных звуковых соответствий (иногда весьма далеких *от звукового тождества) в словах, сходных по значению*, а именно значения слов в данной ситуации и неизвестны. Звуковые соответствия, выведенные на заведомо заимствованных словах, в частности, на адаптированных этрусками именах греческих мифологических персонажей, могут рассматриваться лишь как вспомогательная информация. Поэтому прав И. Фридрих: «С психологической точки зрения вполне понятно, когда этрусковеды особенно охотно прислушиваются к созвучиям и пытаются интерпретировать незнакомые слова этимологически, при помощи сходных по звуковому составу слов известных языков; *однако оттого, что причины таких попыток понятны, сами эти* попытки не перестают быть методически порочными» (Фридрих, 1961, с. 156). Яркой иллюстрацией этой мысли служит анекдот о слуге в русском посольстве в Париже, который, подавая гостье манто, сказал: «Ваш салоп». Дама поняла это как "Vache salope" ("грязная корова" или "распутная корова") и устроила истерику. Между тем, оба были носителями достоверно родственных индоевропейских языков. Легко представить себе, как далеко можно зайти в интерпретации слов по созвучию, если к тому же и родство языков не доказано.

В отличие от этимологического метода, *который является* чисто лингвистическим, комбинаторный включает в себя как дешифровку «изнутри», т. е. анализ комбинаций языковых элементов, так и дешифровку «извне», т. е. привлечение данных самого разного характера: место находки надписи, предмет, на который она нанесена, сопровождающие ее изображения, сведения о греках, италиках и римлянах, современниках и соседях этрусков, с которыми они имели много общего в культуре. Именно применению комбинаторного метода мы обязаны выяснением семантики большинства тех этрусских слов, которые к настоящему времени считаются известными.

Внешние опоры вообще очень важны в истории каждой конкретной дешифровки. Особое место среди них занимает Розеттский камень, послуживший толчком для раскрытия Ж. Ф. Шампольоном древнеегипетского языка и письменности. Эта дешифровка является классическим образцом решения задачи третьего типа. В наших примерах мы моделировали ее текстом на неизвестном нам армянском языке, записанным неизвестной нам армянской графикой.

Но стоявшая перед египтологами задача была намного сложнее. Они имели дело не с алфавитом, в котором большин-

ство знаков обозначает только один звук, а с иероглификой. В этой системе письма номинантом знака может быть: а) слово или морфема, б) согласный или группа согласных, в) информация о значении соседних слов. Один и тот же знак может выполнять то одну, то другую функцию в зависимости от окружения. Например, знак, который передает слово *пер*, что значит «дом», изображает план простейшего дома: стены и вход. Этот же знак может обозначать слог *пер*, точнее, согласные *п* и *р* и неизвестный гласный между ними. Кроме того, он может вообще не произноситься в тексте, но указывать на то, что соседнее слово является обозначением постройки. В этом качестве он является детерминативом. Детерминативы были очень важны, они указывали, следует читать и понимать соседний знак фонетически или как логограмму.

Розеттский камень, найденный французами в 1799 г. во время египетской кампании Бонапарта, представляет собой билингву, т. е. надпись идентичного содержания на двух языках: египетском и греческом. Строго говоря, надпись состояла из трех частей, так как египетский текст был записан, во-первых, иероглифами на древнеегипетском языке, во-вторых, демотическим письмом (нечто вроде упрощенной скорописи) на более позднем новоегипетском. Отправной точкой дешифровки послужило имя Птолемея, обведенное картушем (овальной рамкой). Вслед за своими предшественниками, Шампольон принял гипотезу, что картуш содержит в себе царское имя. Значит, составляющие его знаки обозначают не понятия, а звуки. В поисках номинантов фонетических знаков он обратился к коптскому языку, далекому потомку древнеегипетского, используемому как культовый язык египетских христиан. Опора на билингву и сочетание этимологического и комбинаторного метода позволили Ж. Ф. Шампольону представить египетское письмо как систему идеограмм, фонетических знаков и определителей, а позже дать описание грамматики и лексики древнеегипетского языка.

Говоря о методах дешифровки, уместно коснуться вопроса об интуиции. Так или иначе решать его для себя приходится каждому, кто берется за этот род деятельности.

И. М. Дьяконов подчеркивает «доминирующую роль интуиции» и утверждает, что «человек, обладающий хорошей интуицией, в состоянии уловить структуру текста и указать путь дешифровки, тогда как эрудит, не обладающий интуицией, этого сделать не в состоянии» (Дьяконов, 1976, с. 14). Ему же принадлежит удачное, на наш взгляд, определение интуиции как «способности подсознательно идентифицировать существенные признаки и связи» (Дьконов, 1976, с. 12).

Наше мнение по этому вопросу можно сформулировать так.

1. Акт интуиции, во всяком случае на данном этапе развития человечества, не дан в эксперименте, т. е. мы не можем,

воспроизведя условия, в которых он имел место, добиться его повторения. Интуитивно полученное знание является для ученого таким же непредсказуемым подарком судьбы, как метеоритное железо для народов бронзового века. Между тем, как справедливо утверждает Б. В. Сухотин, «рассматривая задачу дешифровки вообще, мы не можем рассчитывать ни на какие благоприятные обстоятельства» (Сухотин, 1963, с. 75).

2. Интуитивно полученное знание может быть ложным.

3. Мы не отрицаем, что интуитивно полученное знание может быть истинным и путем прозрения человек способен «увидеть» закономерности, реально присутствующие в тексте или даже в языковой системе, *породившей этот текст*. Вопрос состоит в том, является ли таким способом «увиденное» фактом науки. С нашей точки зрения, не является до тех пор, пока не построена вполне рациональная «лестница» умозаключений, которая подвела бы каждого специалиста (а в принципе, каждого желающего) к завоеванной таким образом «крепости». Следовательно, с интуитивно выдвинутой гипотезой поступают так же, как и с гипотезами, выдвинутыми иным путем: ищут *эмпирические данные, подтверждающие или* опровергающие ее, и на основании этих данных строят логические выводы.

Вопрос об интуиции подводит нас к более широкому вопросу о соотношении субъективного и объективного факторов в дешифровке. Всякая конкретная дешифровка имеет в качестве объекта тексты с их особенностями (например, бо́льшая часть этрусских текстов — эпитафии и посвящения, а крито-микенских текстов — хозяйственные записи), а в качестве субъекта — того или иного человека с его талантами и слабостями, знанием или незнанием тех или иных языков, наличием или отсутствием интуиции и т. д.

Но за всем *этим многообразием стоят закономерности*, представляющие собой предмет общей дешифровки, «деятельности, направленной на создание методов распознавания явления языка в текстах на языках, предполагаемых неизвестными, методов, опирающихся исключительно на свойства сочетания элементов текста» (Сухотин, 1984, с. 3—4). Эта дисциплина формируется на слиянии нескольких тенденций. Во-первых, представители частных дешифровок предлагают методы, *позволяющие решать* класс задач, более широкий, чем те, для которых они первоначально разрабатывались. Таков, например, предложенный Ю. В. Кнорозовым метод различения служебных и знаменательных морфем. Во-вторых, структурная и математическая лингвистика рассматривает способы исследования текста без привлечения значения, т. е. искусственно моделирует ситуацию, в какой оказывается дешифровщик. Таковы алгоритмы морфемного членения З. Харриса, статистико-комбинаторный метод Н. Д. Андреева, дешифровочные алгоритмы Б. В. Сухотина, дистрибутивно-статистический анализ А. Я. Шайке-

вича, индуктивная реконструкция М. К. Тимофеевой. В-третьих, обобщают свой опыт военные дешифровщики и специалисты по теории информации (К. Шеннон).

В качестве иллюстрации приведем два алгоритма, решающие сходные задачи: алгоритм морфемной сегментации высказывания, разработанный представителем дескриптивной лингвистики З. Харрисом, и алгоритм отделения основы от аффиксальной части слова, предложенный нами для конкретных целей интерпретации этрусского языка.

Алгоритм З. Харриса. Берется тестовое высказывание U, состоящее из k фонем. Ищутся другие высказывания, начинающиеся с той же фонемы, что и U, и подсчитывается, сколько разных фонем встречается после первой фонемы высказывания U. Они называются ее «преемниками» ("successors"). Затем подбираются высказывания и подсчитывается число «преемников» для последовательности из двух первых фонем тестового высказывания. В общем виде для каждой n-фонемной ($n=1, 2, \ldots, k$) последовательности, составляющей начало тестового высказывания, подсчитывается N — число возможных «преемников». N, таким образом, является некоторой функцией от n. Тестовое высказывание сегментируется в тех местах, т. е. после тех n-фонемных начальных последовательностей, для которых $N=f(n)$ имеет локальный максимум. Аналогично подсчитывается N_1 — число «предшественников» ("predecessors"), т. е. фонем, которые встречаются в других высказываниях перед каждой n-фонемной последовательностью, составляющей конец высказывания U. Тестовое высказывание U сегментируется в тех местах, т. е. перед теми n-фонемными финальными последовательностями, для которых $N_1=f_1(n)$ имеет локальный максимум. Вторая сегментация позволяет уточнить результаты первой.

Опишем наш алгоритм наращения неслучайных буквосочетаний, призванный провести границу между аффиксами и основой. Этрусскому языку свойственна постфиксация, поэтому наращение происходит с конца слова и для работы необходим обратный словарь.

Берем некоторый финальный сегмент x, состоящий из n букв ($n=1, 2, 3, 4, \ldots$), и выясняем, какой процент словоформ, кончающихся на x, кончается на Ax, Cx, Θx, ..., Ψx. Обозначаем эти величины, т. е. относительные частоты буквосочетаний в финальной позиции, как $Ч_{Ax}$, $Ч_{Cx}$ и т. д. Сравниваем относительную частоту буквосочетания Ax с эталонной частотой буквы A (обозначим ее $Ч_A$). Если $Ч_{Ax} \leqslant Ч_A$, то сочетание Ax не является характерным и постоянным для этрусского языка. Если же $Ч_{Ax}$ существенно больше, чем $Ч_A$, то мы вправе предположить, что это не случайно и что сегмент Ax морфологически релевантен.

Так же поступаем с буквосочетаниями Cx, Θx, ..., Ψx. Выделяем те из них, у которых относительная частота в финальной позиции существенно превышает эталонную частоту соответствующей буквы. Принимаем выделенные сегменты снова за x и с каждым из них повторяем описанный цикл.

Как мы видим, структурная лингвистика и дешифровка имеют много точек соприкосновения благодаря лежащим в их основе общим закономерностям строения языка. Надо иметь в виду, что, несмотря на применение ЭВМ, общая дешифровка является не прикладной, а теоретической дисциплиной. Наиболее последовательно это положение отстаивает Б. В. Сухотин. Его цель — не получение результатов, полезных для интерпретации какого-либо неизвестного языка, а отработка методики. Не случайно бо́льшая часть работ по исследованию языка дешифровочными методами проведена им на русском материале.

Свойственные Б. В. Сухотину предельное удаление дешифровки от прикладных задач и игнорирование любой внетекстовой информации вызвало резонное замечание А. М. Кондратова: «Первой (и основной!) задачей дешифровки неизвестного текста является именно его дешифровка» (Кондратов, Шеворошкин, 1970). В. В. Шеворошкин выразил сходную мысль еще определеннее: «Необходимо иметь в виду, что целью дешифровки является не разработка методики процесса исследования, а конечный результат — понимание текста» (Шеворошкин, 1965, с. 203). Однако внесем уточнение: цель — не просто понимание текста (таких пониманий может быть много, в том числе и внутренне непротиворечивых), а правильное его понимание, т. е. такое, которое доказано или, еще лучше, подверглось проверке на другом материале. Вопрос о том, как получить правильное понимание и как отличить его от других, и есть, собственно, вопрос метода.

Поэтому общая дешифровка, вырабатывающая методы, которые обеспечивают возможность интерпретации текстов на произвольных языках, не только имеет право на существование как раздел общего языкознания, но и несомненно полезна для частных дешифровок.

Литература

Андреев Н. Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л., 1967.

Добльхофер Э. Знаки и чудеса. М., 1963.

Дьяконов И. М. О методах дешифровки древних текстов // Тайны древних письмен: Проблемы дешифровки: Сб. статей. М., 1976.

Кнорозов Ю. В. 1) Неизвестные тексты // Забытые системы письма. М., 1982; 2) Письменность индейцев майя. М.; Л., 1963.

Кондратов А. М., Шеворошкин В. В. Когда молчат письмена: Загадки древней Эгеиды. М., 1970.

Савенкова Е. Д. Алгоритм выделения этрусских морфемных сегментов // Романское и германское языкознание. Вып. 1. Минск, 1979.

Сухотин Б. В. 1) Алгоритм лингвистической дешифровки // Проблемы структурной лингвистики / Отв. ред. С. К. Шаумян. М., 1963; 2) Оптимизационные методы исследования языка. М., 1976; 3) Выделение морфем в текстах без пробелов между словами. М., 1984.

Шеворошкин В. В. Исследования по дешифровке карийских надписей. М., 1965.

Шеннон К. Теория связи в секретных системах // Работы по теории информации и кибернетике / Пер. с англ. М., 1963.

Фридрих И. Дешифровка забытых письменностей и языков. М., 1961.

Aalto P. Notes on Methods of Decipherment of Unknown Languages // Studia Orientalia. XI. 4. Helsingfors, 1945.

Harris Z. S. From Phoneme to Morpheme // Language. 1955. Vol. 31, № 2. P. 190—222.

СОДЕРЖАНИЕ

От редактора (А. С. Герд) 3
Предмет и основные направления прикладной лингвистики (А. С. Герд) 5
Введение в теорию письма (Л. Р. Зиндер) 15
Орфография (В. Ф. Иванова) 25
Орфоэпия и вопросы культуры речи (Л. А. Вербицкая) . . . 34
Методика обучения русскому языку как родному в средней школе (В. Г. Ветвицкий) 54
Автоматизированные обучающие системы (И. В. Жарков) . . . 59
Специальный текст как предмет прикладного языкознания (А. С. Герд) 68
Искусственный интеллект (И. П. Панков) 91
Моделирование языка (М. И. Откупщикова) 100
Фонетические аспекты прикладной лингвистики (Л. В. Бондарко) . 113
Моделирование синтаксиса естественного языка (В. Д. Буторов) . 142
Моделирование семантики предложения (В. В. Богданов) . . . 161
Методы статистического моделирования в языкознании (Г. Я. Мартыненко) 201
Статистическая обработка экспериментальных данных (А. С. Штерн) 224
Сетевое представление лингвистических данных (И. В. Азарова) . 234
Психолингвистика (Т. И. Зубкова) 245
Лингвистическая прагматика и ее прикладные аспекты (В. В. Богданов) 268
Прикладное терминоведение и его направления (В. М. Лейчик) . 276
Научно-техническая лексикография (А. С. Герд) 287
Учебная лексикография (М. А. Марусенко) 308
Автоматизация в лексикографии (Л. Н. Беляева, А. С. Герд, И. И. Убин) 318
Информационно-поисковые системы (И. П. Панков, В. П. Захаров) 334
Автоматический (машинный) перевод (Л. Н. Беляева, М. И. Откупщикова) 360
Реферирование (В. В. Богданов) 389
Лингвистическое обеспечение автоматизированного издательского процесса (З. В. Партыко) 399
Лингвистика в картографии (Р. Ю. Кобрин) 408

Стилеметрия (Г. Я. Мартыненко, С. В. Чебанов) 420
Сложность синтаксических структур и стилистическая диагностика (Г. Я. Мартыненко) 435
Компьютерные аспекты стилеметрии (О. Н. Гринбаум) . . . 451
Атрибуция анонимных и псевдонимных текстов методами прикладной лингвистики (М. А. Марусенко) 466
Структурное математическое и прикладное стиховедение (метрика и ритмика) (М. А. Красноперова) 480
Автоматизация обработки древних текстов (С. А. Аверина, Е. Л. Алексеева, А. С. Герд) 509
Дешифровка (Е. Д. Савенкова) 514